UBLICATIONS DE L'INSTITUT DE STATISTIQUE
DE L'UNIVERSITÉ DE PARIS

MÉMOIRES ET CONFÉRENCES SUR LE CALCUL DES PROBABILITÉS,
LA STATISTIQUE THÉORIQUE ET APPLIQUÉE, L'ÉCONOMÉTRIE

Branislav V. IVANOVITCH

SUR LA DISCRIMINATION DES ENSEMBLES STATISTIQUES

VOL. III - FASCICULE 4 - 1954

PARIS
11, Rue Pierre Curie

Toute la correspondance relative aux publications
doit être envoyée à l'adresse :

INSTITUT DE STATISTIQUE DE L'UNIVERSITE DE PARIS

INSTITUT HENRI POINCARÉ - 11, Rue Pierre Curie - PARIS (Ve)

Les manuscrits doivent être envoyés à *M. Daniel DUGUÉ,*
à l'adresse précédente.

Abonnements : Pour la France 1.200 francs français
Pour l'Etranger 1.500 francs

Vente au numéro : *(fascicule de 50 pages environ)*
Pour la France 350 francs français
Pour l'Etranger 400 francs

[illegible]

[illegible]

SUR LA CONSERVATION

DES ENSEMBLES STATISTIQUES

VOL. III — FASCICULE [illegible] — [illegible]

[illegible]



UBLICATIONS DE L'INSTITUT DE STATISTIQUE
DE L'UNIVERSITÉ DE PARIS

MÉMOIRES ET CONFÉRENCES SUR LE CALCUL DES PROBABILITÉS
LA STATISTIQUE THÉORIQUE ET APPLIQUÉE, L'ÉCONOMÉTRIE

Branislav V. IVANOVITCH

SUR LA DISCRIMINATION DES ENSEMBLES STATISTIQUES

VOL. III - FASCICULE 4 - 1954

PARIS
11, Rue Pierre Curie

Étienne HALPHEN

1911-1954

Etienne HALPHEN était né le 27 mai 1911 à Bordeaux, où son père était alors professeur à la Faculté des Lettres. Son grand-père, Georges-Henri HALPHEN, éminent mathématicien, mort à quarante-quatre ans, avait laissé une oeuvre considérable et variée. (Théorie des courbes algébriques, Traité sur les fonctions élliptiques, etc.)

Etienne HALPHEN manifesta de bonne heure des dons exceptionnels pour la science mathématique : à dix-sept ans, élève de mathématiques élémentaires au Lycée de Bordeaux, il donna "Un Théorème sur les quadriques analogue à celui de Chasles sur les coniques", qui fit grande impression parmi les géomètres de l'époque. (C.R. de l'Académie des Sciences du 18 mars 1929).

Reçu en 1930 aux concours d'entrée à l'Ecole Polytechnique et à l'Ecole Normale Supérieure, il choisit cette dernière où il devint agrégé de mathématiques en 1933. Professeur au Lycée de Sens durant un an, sa santé précaire le contraignit vite à abandonner l'enseignement.

En 1936, il entra à la Recherche Scientifique. Durant quatre années, il allait se consacrer à la Statistique mathématique, dont les méthodes si fécondes venaient d'être introduites dans notre pays par M. le Professeur Georges DARMOIS. Les travaux qu'il poursuivit jusqu'en 1940 concernent tant la théorie que les applications. Sur le plan théorique, il étudia notamment des procédés d'estimation et mis au point sa "fonction caractéristique intrinsèque", outil fécond et parfaitement original pour l'étude des distributions de probabilité. Ce travail fait d'Etienne HALPHEN un précurseur : les idées qui le guidaient sont celles-là mêmes que retrouvent aujourd'hui les mathématiciens qui construisent la Théorie de l'Information. En matière d'application, il s'occupa notamment de problèmes de biométrie, puis, en 1939-40, de problèmes intéressant la Défense Nationale.

Puis ce fut l'occupation. Parmi tant d'autres, Etienne HALPHEN et sa famille devaient en souffrir. C'est pourtant dans cette période troublée qu'il commença à s'occuper d'hydrologie et d'économie électrique. Des circonstances fortuites l'avaient mis en contact à Grenoble avec M. Pierre MASSE qui dirigeait à l'époque plusieurs sociétés hydroélectriques et qui se préoccupait de mettre sur pied des méthodes susceptibles de contribuer à une solution rationnelle des problèmes d'exploitation et d'équipement. Pour remplir cette tâche, des connaissances économiques ne suffisaient pas. Le caractère aléatoire des débits de nos rivières posait de nouveaux

problèmes. Il fallait jeter les bases d'une économie de l'incertain. Pour y parvenir, une formation statistique solide était indispensable. Etienne HALPHEN possédait cette formation. Mais il possédait bien autre chose : une aptitude à saisir dès la première rencontre la pensée de son interlocuteur, une ouverture d'esprit grâce à laquelle aucun problème ne lui demeurait étranger, un sens du concret qui mettait ses plus hautes spéculations au service de l'efficacité. Invité par M. Pierre MASSÉ à travailler pour la Société Hydrotechnique de France, il eut à étudier les problèmes les plus divers ; modèle hydraulique, prévision des crues et des étiages, valeur comparée des divers équipements ... Pour un statisticien les données de base manquaient terriblement et des méthodes adéquates étaient à créer de toutes pièces. Ce fut là désormais une de ses tâches essentielles.

Lors de la création d'Electricité de France, Etienne HALPHEN et ses collaborateurs furent rattachés au Service des Etudes et Recherches Hydrauliques dirigé par M. NI ZERY, qui fut enlevé à notre amitié il y a seulement quelques mois. C'est là qu'il continuait, comme ingénieur-conseil, à nous apporter la contribution irremplaçable d'une intelligence hors série.

Il n'est pas possible de retracer dans le détail les travaux menés par Etienne HALPHEN ou sous sa direction en matière d'hydrologie. Quelques-uns sont bien connus de la Société Hydrotechnique de France. Mentionnons les mémoires publiés dans l'Annuaire Hydrologique en 1945 et 1946 : y furent exposés quelques travaux statistiques qui contribuèrent à la préparation du premier Plan d'équipement. Mentionnons encore la communication qui fut faite en 1949 sur les lois de probabilité des débits des cours d'eau. A ce propos, nous voudrions dire combien Etienne HALPHEN prenait à coeur ce problème des formes analytiques des lois de probabilité. Il y a plus de dix ans qu'il avait montré l'intérêt et la commodité d'emploi de celle qu'il avait appelée "loi harmonique". Pour les besoins de l'hydrologie, avaient été introduites des formes plus générales qui garderont le nom de "lois Halphen" des types A et B. Leur auteur observait il y a quelques mois ; "De telles lois manquaient en Statistique. C'est avec un certain étonnement que nous constatons ce fait ; cinquante ans après les remarquables travaux de Karl Pearson, nul ne semble d'être avisé que les lois de Person, si précieuses qu'elles aient été, sont très insuffisantes."

L'étude des lois B avait conduit à envisager sous un jour nouveau une importante classe de fonctions ; il s'agit des fonctions hypergéométriques confluentes connues sous le nom de "fonctions de Hermite" qu'il préférait appeler "fonctions factorielles". Elles lui avaient donné l'occasion de découvrir une identité fonctionnelle bien remarquable, qui fait éprouver à tout mathématicien un plaisir et une surprise intenses :

$$\int_0^\infty e^{-z^2}(\cos xz + \sin xz)\frac{dz}{\sqrt{z}} \equiv \int_0^\infty e^{-(z-\frac{x}{2})^2}\frac{dz}{\sqrt{z}}$$

quel que soit x, réel ou complexe. Ce qui est frappant, c'est le fait que l'imaginaire ne figure dans aucun des deux membres de cette identité reliant l'exponentielle et les fonctions circulaires. Pour Etienne HALPHEN, cette identité n'est qu'un cas particulier d'une "formule de réflexion" reliant deux catégories de fonctions factorielles.

Signalons encore parmi ses travaux les plus récents, une utilisation nouvelle de l'analyse harmonique dans l'étude des débits d'un cours d'eau. L'essentiel de la méthode a été indiqué dans une communication au Congrés de l'Union Géophysique Internationale que s'est tenu à Rome en sep-

tembre. Son auteur avait grande confiance dans cette méthode d'analyse. Pourtant les bases théoriques restent à étayer, et c'est un difficile problème. Mais les résultats concrets sont déjà assez remarquables pour qu'on accorde crédit à la méthode. Ainsi procède souvent le statisticien.

Encore un mot sur les méthodes de travail d'Etienne HALPHEN et sur sa forme d'esprit. Deux traits fondamentaux le caractérisent ; C'est d'abord l'habitude et le besoin de repenser intégralement toute théorie, mathématique ou autre. A toute occasion il savait dégager des vues nouvelles, pratiquement toujours fécondes. Ensuite, il croyait au primat de l'empirisme : la théorie n'est qu'un outil qu'il faut adapter au mieux. Ajoutez à cela une forte dose d'intuition, et une honneteté intellectuelle scrupuleuse. Ne voilà-t-il pas les qualités qui définissent le statisticien idéal ?

Son vieux rêve était que l'Ecole française de statisticiens, fût une digne émule des écoles anglo-saxonnes. Ceux qui ont eu la chance incomparable de travailler à ses côtés savent que, par les travaux nombreux qu'il nous laisse à exploiter ou à poursuivre, par sa façon originale d'envisager les problèmes, par la formation qu'il a donnée à quelques-uns, Etienne HALPHEN occupera une place de choix dans la statistique française.

G. MORLAT

SUR LA DISCRIMINATION DES ENSEMBLES STATISTIQUES

par

Branislav V. IVANOVITCH

1. INTRODUCTION

En mesurant un caractère commun des éléments de deux populations π_1 et π_2 nous obtenons les lois de probabilité $f_1(x)$ resp. $f_2(x)$ qui peuvent renfermer un domaine commun dans l'espace R^1 (fig. 1). La distinction entre deux populations en vertu de la mesure d'un seul caractère s'effectue aujourd'hui selon un procédé standardisé, dont la théorie est très développée. Cependant, les conclusions que l'on obtient sont souvent insuffisantes et on est obligé d'entreprendre les mesures de nouveaux caractères communs des éléments appartenant à deux populations pour avoir des renseignements supplémentaires dans le but d'arriver à des conclusions plus précises.

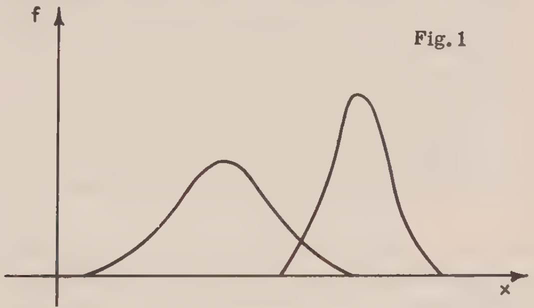


Fig. 1

En tenant compte de ces renseignements supplémentaires, on a été obligé d'élargir la théorie statistique. Ainsi dans le cas de deux ou plusieurs variables aléatoires. En outre, il existe déjà plusieurs méthodes pratiques qui élargissent les théories statistiques pour le cas de n variables aléatoires. Presque toutes ces méthodes sont basées sur la distribution commune des variances et covariances d'un échantillon de la population dont les caractères observés suivent les lois de Laplace. Wishart (1,2) et Bartlett (2) ont donné la forme analytique de cette distribution. Deux ans plus tard Hotelling (3) a réussi à donner la généralisation de la loi de "Student" (4). Enfin, Barlett (5,7), Wilks (6), Hotelling (8), Plackett (9), Rao (10,11) ont publié plusieurs tests d'hypothèse servant à examiner : a) l'indépendance de plusieurs groupes de variables aléatoires ; b) l'égalité des moyennes arithmétiques de n variables aléatoires dépendantes d'une population normale ; c) l'égalité des matrices de dispersion de deux populations normales.

R. A. Fisher (12, 13, 14, 15) a indiqué une nouvelle direction dans l'étude de la séparation de deux ou plusieurs populations en y introduisant la méthode des fonctions de discrimination. Il a montré qu'avec une suite de mesures de plusieurs caractères on peut former une fonction de discrimination, linéaire selon les caractères mesurés, ayant la propriété de séparer deux populations mieux que n'importe quelle autre fonction linéaire. Ainsi on a réduit le problème au cas d'une seule variable aléatoire, représentée sous la forme d'une expression linéaire suivant les caractères observés. Les coefficients ont été déterminés de telle manière que la valeur de cette expression atteigne son maximum. Cette méthode était appliquée jusqu'à présent à la technologie (19), à la sélection des plantes (18), aux sciences des finances (21), à la taxonomie (12), à la crâniométrie (16), à la psychologie (20,24), à la zoologie (25) etc..

Le but de notre travail était de présenter quelques nouveaux problèmes de discrimination et de donner leurs solutions.

Dans le premier chapitre nous avons exposé la théorie connue de la discrimination de deux populations en ajoutant toutefois quelques remarques. Nous avons d'abord présenté une interprétation géométrique de la fonction de discrimination de Fischer dans le cas où les deux populations sont normalement distribuées et ont les mêmes matrices de dispersion. Ensuite, nous avons montré comment on peut déterminer un plan de discrimination dans le cas où les deux populations sont normales et ont des axes parallèles. Lorsque les deux populations sont distribuées d'une façon quelconque, Welch est arrivé à déterminer la surface de discrimination dont la forme est en général compliquée. C'est pourquoi nous avons essayé, tout en nous servant des mêmes principes que ceux de Welch, de former une nouvelle surface de discrimination qui serait toujours un plan. Dans le cas où les deux populations sont normales et ont les mêmes matrices de dispersion, les deux surfaces se réduiront au plan de discrimination de Fisher.

Dans le deuxième chapitre nous avons élargi la théorie de discrimination en considérant le cas où les caractères d'un ou de plusieurs éléments se sont transformés sous l'influence de certaines causes. Si, par exemple, l'élément qu'il faut classer par rapport à deux populations se transforme de cette façon que les valeurs de ses caractères changent proportionnellement, il est évident que les méthodes précédentes, basées sur la différence des moyennes arithmétiques des caractères, ne correspondront plus à la nature du problème. Pour la discrimination de deux populations ne sont plus essentielles seulement les différences des moyennes arithmétiques, mais aussi leurs proportions. Ces considérations nous ont conduit à construire une fonction, dite "fonction de disproportion", qui s'adapterait mieux à la nature du problème mentionné. Dans le cas où les valeurs des caractères de tous les éléments changent proportionnellement, d'après une loi connue de transformation, nous montrerons comment il est possible de résoudre le problème de la discrimination au moyen de la méthode de Welch.

Lorsque la surface de discrimination est fixée d'avance et la forme mathématique des lois de probabilité de deux populations, contenant les mêmes paramètres inconnus, est connue, nous montrerons dans le troisième chapitre comment il faut déterminer les valeurs des paramétres inconnus afin que les distributions des populations soient établies de telle sorte que la surface de discrimination les sépare aussi bien que possible.

Enfin, dans le quatrième chapitre de ce travail, on a observé le cas où nous connaissons la loi de probabilité de la population totale $\pi = \pi_1 + \pi_2$,

sans connaitre celle de π_1 et π_2. Afin de pouvoir entreprendre la classification des éléments particuliers de la population π par rapport aux populations π_1 et π_2, il est nécessaire de déterminer préalablement les lois de probabilité de ces populations. Dans le cas où il est possible d'exprimer la loi de probabilité de la population π_2 au moyen de celle de π_1 par la transformation linéaire suivante.

$$f_2(x) = \int_{-\infty}^{+\infty} f_1(\xi) K(x,\xi) d\xi ,$$

nous montrerons que le problème se réduit alors à résoudre une équation intégrale de Fredholm de seconde espèce dont la solution sera $f_1(x)$.

Si l'on connait une suite de couples $f_{1i}(x)$, $f_{2i}(x)$, $i = 1,2...N$, nous avons montré dans le même chapitre comment il est possible de trouver une approximation en moyenne du noyau $k(x,\xi)$. Une telle détermination de $k(x,\xi)$ nous permettra d'établir une dépendance statistique entre les lois de probabilité $f_1(x)$ et $f_2(x)$. Cette dépendance peut être plus ou moins forte et nous établirons un coefficient représentant la mesure de cette dépendance.

Je suis heureux de pouvoir exprimer ici à M. le Professeur Georges Darmois toute ma reconnaissance pour son aide constante, ses encouragements et ses conseils qui m'ont été si utiles au cours de ce travail.

2. PLAN DE DISCRIMINATION

Classer un élément, selon ses caractères mesurés, par rapport à deux populations données π_1 et π_2, qui s'étendent dans un espace à k dimensions R^k, représente un des problèmes principaux que se pose l'analyse statistique à plusieurs dimensions. Cette classification peut être faite en séparant l'espace R^k en deux régions R_1^k et R_2^k qui s'excluent mutuellement ce qui nous permettrait d'accepter l'hypothèse selon laquelle l'élément observé appartient à la population π_i s'il se trouve dans la région R_i^k (i = 1,2). Il est évident que cette séparation de l'espace R^k doit être faite aussi bien que possible, afin que l'erreur de classification soit réduite au minimum.

Supposons d'abord que les populations π_1 et π_2 sont normales non-singulières et qu'elle ont les mêmes matrices de dispersion. Soient m_x^i et m_y^i les moyennes arithmétiques de i-ème caractère des population π_1 et π_2,

$$\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1k} \\ w_{12} & w_{22} & \cdots & w_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ w_{1k} & w_{2k} & \cdots & w_{kk} \end{bmatrix}$$

la matrice de dispersion commune et

$$f_1(x_1, x_2, \ldots, x_k) = \frac{1}{(2\pi)^{K/2}\sqrt{W}} e^{-\frac{1}{2}\sum_{\zeta}\sum_{\nu} w^{\zeta\nu}(x_\zeta - m_x^\zeta)(x_\nu - m_x^\nu)}$$

et

$$f_2(x_1, x_2, \ldots, x_K) = \frac{1}{(2\pi)^{K/2}\sqrt{W}} e^{-\frac{1}{2}\sum_{\zeta}\sum_{\nu} w^{\zeta\nu}(x_\zeta - m_y^\zeta)(x_\nu - m_y^\nu)}$$

les lois de probabilité de ces deux populations où W représente la valeur du déterminant de la matrice de dispersion et $w^{\zeta\nu}$ les éléments de la matrice réciproque. Vu la supposition selon laquelle les populations sont normales non-singulières, il résulte qu'il n'existe aucun hyperplan dans l'espace R^k pouvant contenir la masse totale de la population π_i (i ÷ 1,2).

Soient $\xi_1, \ldots, \xi_k$ les mesures des caractères d'un élément tiré de la population π. Nous savons à priori que cette population ne saurait être que π_1 ou π_2.

Afin de donner une base de classification de l'élément $\xi(\xi_1, \ldots, \xi_k)$. posons une surface $\varphi(x_1, \ldots, x_k) = 0$ qui partagerait l'espace R^k un deux régions R_1^k et R_2^k. Pour simplifier, supposons que la surface est un hyperplan, c'est-à-dire.

$$(2.1) \qquad \varphi(x_1, \ldots, x_k) \equiv \sum_{\nu=1}^{k} (x_\nu - M_\nu)\cdot\lambda_\nu = 0$$

et déterminons les coefficients λ_ν de sorte que ce plan de discrimination sépare aussi bien que possible π_1 et π_2. Considérons un élément $x(x_1,..x_k)$ de la pupulation π_1. La distance de celui-ci au plan (2.1) est

$$d = \sum_{\nu=1}^{k} (x_\nu - M_\nu)\lambda_\nu .$$

La moyenne arithmétique de ces distances relative à la masse totale de la population π_1 est donnée par la formule

$$E_1(d) = E_1\left\{\sum_{\nu=1}^{k} (x_\nu - M_\nu)\lambda_\nu\right\},$$

c'est-à-dire

$$E_1(d) = \sum_{\nu=1}^{k} (m_x^\nu - M_\nu)\lambda_\nu .$$

On a également

$$E_2(d) = \sum_{\nu=1}^{k} (m_y^\nu - M_\nu)\lambda_\nu$$

pour la moyenne arithmétique de la distance des éléments au plan (2.1) relative à la population π_2. La variance de la distance pour la population π_i aura la forme

$$\text{Var.}_i(d) = E_i\left[d - E_i(d)\right]^2 = E_i\left[\sum_{\nu=1}^{k}\left\{x_\nu - E_i(x_\nu)\right\}\lambda_\nu\right]^2 ,$$

c'est-à-dire

$$\text{Var.}_i(d) = \sum_{\zeta=1}^{k}\sum_{\nu=1}^{k} \lambda_\zeta\lambda_\nu W_{\zeta\nu}^{(i)} , \quad i = 1,2 .$$

Vu l'identité des matrices de dispersion, il devient

$$\text{Var.}_1(d) = \text{Var.}_2(d) = \text{Var.}(d) = \sum_{\zeta=1}^{k}\sum_{\nu=1}^{k} \lambda_\zeta\lambda_\nu W_{\zeta\nu} .$$

Dans le but d'établir une mesure commune concernant les distances des éléments des deux masses au plan (2.1), divisons ces distances par la variance commune Var. (d). Les valeurs d/Var. (d) peuvent être positives ou négatives, selon que les éléments correspondants et le centre des coordonnées se trouvent ou ne se trouvent pas du même côté du plan (2.1).

Afin de séparer les populations π_1 et π_2 aussi bien que possible, nous allons déterminer la position du plan de telle sorte que le produit des masses de la population π_i, situées dans la régions R_i^k, par leur distances normées à ce plan soit maximum, tandis que le produit des masses de π_i dans R_j^k $(i \neq j)$ par leurs distances à ce même plan soit minimum. En d'autres termes, la masse totale de π_i dans R_i^k doit être aussi grande et en même temps aussi éloignée que possible du plan de discrimination, tandis que la masse de π_i dans R_j^k doit être aussi petite et aussi proche que possible du même plan. On pourrait alors déterminer les coefficients λ_ν de telle sorte que l'expression

$$\frac{\left\{E_1(d) - E_2(d)\right\}^2}{\text{Var.}(d)} = \frac{\left\{\sum_{\nu=1}^{k}\lambda_\nu(m_x^\nu - m_y^\nu)\right\}^2}{\sum_{\zeta=1}^{k}\sum_{\nu=1}^{k}\lambda_\zeta\lambda_\nu W_{\zeta\nu}}$$

avec la condition $\sum_{\nu=1}^{k} \lambda_\nu^2 = 1$, atteigne son maximum. En désignant

$$d_\nu = m_x^\nu - m_y^\nu , \quad \sum_{\nu=1}^{k}\lambda_\nu d_\nu = D , \quad \sum_\zeta\sum_\nu \lambda_\zeta\lambda_\nu W_{\zeta\nu} = S , \quad \lambda_\nu = \frac{\alpha_\nu}{\sqrt{\sum_{\nu=1}^{k}\alpha_\nu^2}}$$

et en annulant les dérivées partielles par rapport aux α_ν, on obtient le système d'équations.

$$\sum_{\zeta=1}^{k} \alpha_\zeta w_{\zeta\nu} = \frac{S}{D} d_\nu , \qquad \nu = 1, 2, \ldots, k .$$

Vu que S/D figure en facteur, les coefficients cherchés seront proportionnels aux solutions du système linéaire

$$(2.2) \qquad \begin{aligned} w_{11} \lambda_1 + w_{12} \lambda_2 + \ldots\ldots + w_{1k} \lambda_k &= d_1 \\ w_{12} \lambda_1 + w_{22} \lambda_2 + \ldots\ldots + w_{2k} \lambda_k &= d_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ w_{1k} \lambda_1 + w_{2k} \lambda_2 + \ldots\ldots + w_{kk} \lambda_k &= d_{k'} \end{aligned}$$

c'est-à-dire aux valeurs

$$(2.3) \qquad \lambda_\nu = \sum_{\zeta=1}^{k} w^{\zeta\nu} d_\zeta , \qquad \nu = 1, 2, \ldots, k .$$

En utilisant (2.3) on peut écrire D dans la forme suivante

$$D = \sum_{\nu=1}^{k} \lambda_\nu d_\nu = \sum_{\zeta=1}^{k} \sum_{\nu=1}^{k} w^{\zeta\nu} d_\zeta d_\nu .$$

En multipliant la ν-ème équation du système (2.2) par λ_ν et en additionnant toutes les équations du système, nous aurons

$$\sum_{\zeta=1}^{k} \sum_{\nu=1}^{k} w_{\zeta\nu} \lambda_\zeta \lambda_\nu = \sum_{\nu=1}^{k} \lambda_\nu d_\nu ,$$

c'est-à-dire

$$S = D .$$

Par conséquent

$$\text{Max.} \left\{ \frac{D^2}{S} \right\} = \sum_{\zeta=1}^{k} \sum_{\nu=1}^{k} w^{\zeta\nu} d_\zeta d_\nu .$$

Il reste encore à déterminer le point M par lequel passe le plan de discrimination. Etant donné que les variable $x_1, \ldots, x_k$ suivent les lois de Laplace, on peut les exprimer par des variables nouvelles $\xi_1, \ldots, \xi_k$ qui seront indépendantes en suivant également les lois de Laplace. En appliquant le même procédé à ces nouvelles variables, nous obtiendrons les coefficients λ'_ν, tandis que les expressions

$$(2.5) \qquad \frac{d - E_1(d)}{\sqrt{\text{Var.}(d)}} = \frac{\sum_{\nu=1}^{k} (\xi_\nu - m_\xi^\nu) \lambda'_\nu}{\sqrt{\sum_{\nu=1}^{k} \lambda'^2_\nu \sigma^2_{\xi\nu}}}$$

et

$$(2.6) \qquad \frac{d - E_2(d)}{\sqrt{\text{Var.}(d)}} = \frac{-\sum_{\nu=1}^{k} (\xi_\nu - m_\eta^\nu)\lambda'_\nu}{\sqrt{\sum_{\nu=1}^{k} \lambda'^2_\nu \sigma^2_{\eta_\nu}}}$$

où m_ξ^ν , m_η^ν , et σ_{ξ_ν} , σ_{η_ν} sont les paramètres transformés des deux populations, auront les distributions normales N (0,1).

Si nous prenons à présent un élément avec des mesures transformées $\xi_1^\circ , \dots , \xi_k^\circ$ qu'on veut classer, c'est-à-dire déterminer s'il appartient à la population π_1 ou à la population π_2 , nous obtiendrons - en remplaçant ξ_ν par ξ_ν° dans les formules (2.5) et (2.6) - deux valeurs numériques λ_{p_1} et λ_{p_2} qui représentent les écarts normaux correspondant aux probabilités de p_i % (i = 1,2). Si λ_{p_i} dépasse le seuil de signification, on rejettera l'hypothèse que l'élément considéré appartient à la population π_i . Remarquons que le plan de discrimination est parallèle aux plans

$$(2.7) \qquad \sum_{\nu=1}^{k} (\xi_\nu - m_\xi^\nu)\lambda'_\nu = \lambda_{p_1}\sqrt{\sum_{\nu=1}^{k} \lambda'^2_\nu \sigma^2_{\xi_\nu}}$$

et

$$(2.8) \qquad \sum_{\nu=1}^{k} (\xi_\nu - m_\eta^\nu)\lambda'_\nu = \lambda_{p_2}\sqrt{\sum_{\nu=1}^{k} \lambda'^2_\nu \sigma^2_{\eta_\nu}}$$

et déterminons le point M de telle sorte que le plan de discrimination représente le lieu géométrique des point ξ pour lesquels on a $\lambda_{p_1} = -\lambda_{p_2}$, c'est-à-dire

$$\Pr.(\xi \in \pi_1) = \Pr.(\xi \in \pi_2).$$

Il est facile de constater qu'on a dans ce cas

$$\lambda_{p_1} = -\lambda_{p_2} = -\frac{\sum_{\nu=1}^{k} \lambda'_\nu d_\nu}{2\sqrt{\sum_{\nu=1}^{k} \lambda'^2_\nu \sigma^2_\nu}}$$

et

$$M_\nu = \frac{m_\xi^\nu + m_\eta^\nu}{2} .$$

Par conséquent, le plan de discrimination

$$(2.9) \qquad \sum_{\nu=1}^{k} \frac{d_\nu}{\sigma^2_\nu}\left(\xi_\nu - \frac{m_\xi^\nu + m_\eta^\nu}{2}\right) = 0$$

partage l'espace R^k en deux régions R_1^k et R_2^k de telle manière que, si l'élément examiné se trouve dans la région R_j^k , on a

$$\Pr.(\xi \in \pi_i) \leqslant \Phi\left(\frac{1}{2}\sqrt{\sum_{\nu=1}^{k} \frac{(m_\xi^\nu - m_\eta^\nu)^2}{\sigma^2_\nu}}\right) \leqslant \Pr.(\xi \in \pi_j) , \quad i \neq j ,$$

où l'on a posé

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2}\, dt .$$

Si l'on prend $\lambda_{p_1} = -\lambda_{p_2}$ = seuil de signification, les plans parallèles (2.7) et (2.8) partageront l'espace R^k en trois régions $R_1^k - C_1$, $C_1 + C_2$ et $R_2^k - C_2$. Si l'élément examiné $\xi°$ se trouve dans la région $R_i^k - C_i$, on acceptera l'hypothèse qu'il appartient a la population π_i $(i = 1; 2)$ S'il se trouve dans la région C_i, cette hypothèse ne saurait être acceptée et tout ce que nous pouvons dire, c'est qu'il est plus probable qu'il appartient à la population π_i qu'à la population π_j $(i \neq j)$.

L'expression

$$D = \sum_{\nu=1}^{k} \lambda_\nu d_\nu = \sum_{\zeta=1}^{k} \sum_{\nu=1}^{k} w^{\zeta\nu} d_\zeta d_\nu$$

est nommée par P.C.Mahalanobis (22,23,26,33) "distance généralisée". Elle représente la distance maximum à une dimension entre les populations π_1 et π_2. Fisher, ainsi que Mahalanobis, considère cette distance comme une mesure de contraste entre les deux populations. Plus cette distance est grande, plus petite sera l'erreur de classification d'un élément examiné par rapport aux deux populations. Il est naturel de s'y attendre que l'erreur de classification diminue à mesure que le nombre de caractères observés augmente. Afin d'examiner cela, supposons d'abord que nous avons mesuré k caractères dont au moins une des différences $d_\nu = m_x^\nu - m_y^\nu$, $(\nu = 1,2,\ldots,k)$, n'est pas nulle. La valeur correspondante de la distance entre les deux populations sera

$$D_k = \sum_{\nu=1}^{k} {}^{*} \lambda_\nu d_\nu , \tag{2.10}$$

où les coefficients λ_ν sont représentés par la solution du système linéaire

$$\begin{aligned} w_{11}\lambda_1 + \ldots + w_{1k}\lambda_k &= d_1 \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ w_{1k}\lambda_1 + \ldots + w_{kk}\lambda_k &= d_k . \end{aligned} \tag{2.11}$$

Ajoutons à présent r nouveaux caractères que nous pouvons supposer indépendants des k premiers. Nous avons donc en tout k + r caractères et la distance entre les deux populations sera

$$D_{k+r} = \sum_{\nu=1}^{k+r} \lambda_\nu d_\nu .$$

En supposant $d_\nu = 0$ pour $\nu = k+1,\ldots,k+r$, les coefficients λ_ν représentent la solution du système linéaire

$$\begin{bmatrix} w_{11} & \ldots & w_{1k} & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ w_{1k} & \ldots & w_{kk} & 0 & \ldots & 0 \\ 0 & \ldots & 0 & w_{k+1,k+1} & \ldots & w_{k+1,k+r} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & w_{k+1,k+r} & \ldots & w_{k+r,k+r} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \cdot \\ \lambda_k \\ \lambda_{k+1} \\ \cdot \\ \lambda_{k+r} \end{bmatrix} = \begin{bmatrix} d_1 \\ \cdot \\ d_k \\ 0 \\ \cdot \\ 0 \end{bmatrix} \tag{2.12}$$

de k + r équations, qui se décompose en deux, dont le premier se réduit au système (2.11) et le second au système homogène

$$(2.13)\qquad \begin{array}{l} w_{k+1,k+1}\,\lambda_{k+1} + \ldots + w_{k+1,k+r}\,\lambda_{k+r} = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ w_{k+1,k+r}\,\lambda_{k+1} + \ldots + w_{k+r,k+r}\,\lambda_{k+r} = 0 \end{array}$$

Pour cette raison les coefficients $\lambda_1, \ldots, \lambda_k$ conservent les mêmes valeurs et vu la supposition $d_{k+1} = \ldots = d_{k+r} = 0$, on obtient

$$D_{k+r} = D_k .$$

La distance dans ce cas-là n'a pas augmenté bien que nous ayons ajouté de nouveaux caractères et que, si l'on prend par exemple k = 1 et r = 1, on

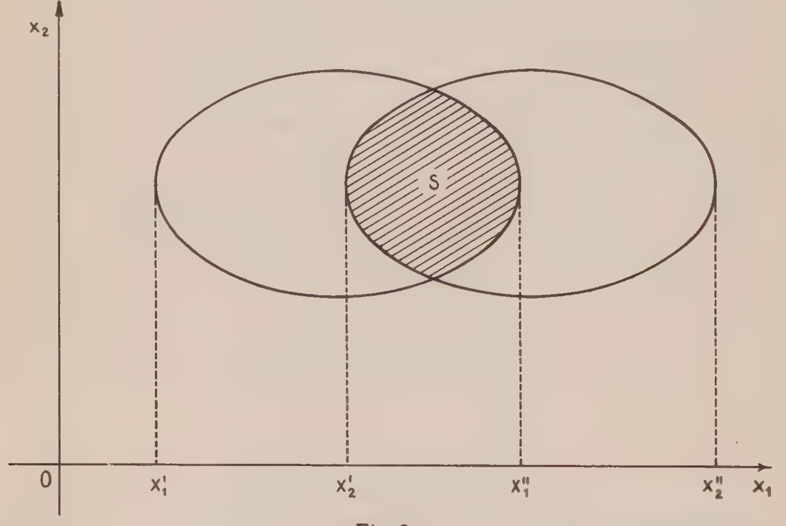


Fig. 2

constate sur la fig. 2 que la masse commune diminue lorsqu'on ajoute un nouveaux caractère, c'est-à-dire

$$\int_{x_1'}^{x_1''} \left[f_1(x_1) + f_2(x_1) \right] dx_1 \geqslant \iint_S \left[f_1(x_1, x_2) + f_2(x_1, x_2) \right] dx_1\, dx_2 .$$

Exemple numérique. - Considérons l'exemple suivant traité par Radhakrishna Rao (27) relatif à la crâniométrie. Au cours du mois d'aout 1939, une société anglaise d'archéologie a fait des fouilles du camp High-down Hill (aux environs de Goring by Sea, Sussex, Angleterre). A cette occasion on a exhumé un squelette humain masculin dans un état relativement bien conservé. D'après les objets trouvés à côté du squelette on a conclu qu'il date environ 500 ans d'avant J.C., c'est-à-dire au commencement de l'âge de fer en Sussex. A cette époque les conquérants "de fer" ont envahi Sussex, la population ayant été mélangée, la question se pose si le squelette trouvé appartenait à un homme de l'âge d'airain ou de l'âge de fer

Afin de résoudre cette question, il est nécessaire de connaitre les distributions des deux populations. Rao a utilisé les données recueillies et publiées sur les squelettes de l'âges d'airain (Morant (30)) et celles de l'âge de fer-romano-britannique (Goodwin et Morant (31)). On a pris les moyennes de sept caractères du crâne suivants ;

L . . . . . diamètre antéro-postérieur

B . . . . . diamètre transverse

B'. . . . . diamètre frontal minimum

S . . . . . courbe sagitalle totale

βQ' . . . . courbe transversale biauviculaire

H'. . . . . hauteur basion-bregma

G'H . . . . hauteur faciale supérieure

Le Tableau 2.1 contient les moyennes arithmétiques des longueurs (exprimées en cm.) de ces caractères pour les deux âges en question, de même que les valeurs de ces mêmes caractères pour le crâne de Highdown. Rao suppose que les populations des deux âges sont normales et que les matrices de dispersion sont identiques à la matrice de dispersion d'une grande population des crânes humains masculins de XVIIe siècle (Hooke (32)).

En appliquant la méthode de χ^2 en vue de tester l'hypothèse H_1, c'est-à-dire si le crâne de Highdown appartenait à l'âge d'airain, on obtiendrait une probabilité de 20 % ce qui veut dire que nous ne saurions rejeter l'hypothèse H_1. Pour cette raison, bien que la même méthode donne une probabilité de plus de 78 % que le crâne examiné appartient à l'âge de fer romano-britannique, la question reste ouverte.

TABLEAU 2.1

caractères	moyennes arithmétiques		crâne de Highdown
	âge d'airain	âge de fer	
L	184,5	188,6	198,2
B	149,9	140,8	148,1
B'	99,7	97,1	101,2
S	376,6	383,5	407,2
βQ'	320,5	316,7	333,9
H'	134,9	137,1	142,0
G'H	69,1	72,4	72,4

En utilisant la méthode de χ^2 nous avons examiné si l'on peut accepter l'hypothèse H_1, à savoir que l'élément considéré appartient à la population π_1, sans tenir compte du fait connu qu'il n'existe qu'une seule hypothèse alternative H_2.

Etant donné que la fonction de discrimination tient compte justement de ce fait, nous pouvons espérer d'obtenir des résultats plus précis. En effet, en appliquant la méthode des fonctions de discrimination, Rao a obtenu les probabilités de 0,84 % et 59,8 %, ce qui veut dire qu'on peut

accepter l'hypothèse que le crâne de Highdown appartenait à l'âge de fer romano-britannique.

3. DÉTERMINATION DU PLAN DE DISCRIMINATION LORSQUE LES POPULATIONS π_1 ET π_2 SONT NORMALES ET LES AXES PRINCIPAUX PARALLÈLES

Soient de nouveau les population π_1 et π_2 normales et les axes principaux parallèles, ce qui peut arriver lorsque par exemple

$$W_{\zeta\nu}^{(1)} \lesseqgtr W_{\zeta\nu}^{(2)}$$
$$W_{\nu\nu}^{(1)} = W_{\nu\nu}^{(2)}$$

Dans ce cas nous pouvons aussi exprimer les variables $x_1, \dots, x_k$ par des variables nouvelles $\xi_1, \dots, \xi_k$ qui seront indépendantes et Laplaciennes. Soient $\sigma_{\nu 1}$ et $\sigma_{\nu 2}$ les écarts-types des populations transformées et désignons $d_\nu = m_\xi^\nu - m_\eta^\nu$. Il s'agit ici de deux fonctions de discrimination

$$D_i = \sum_{\nu=1}^{k} \lambda_{\nu i} d_\nu , \quad i = 1, 2 ,$$

et nous déterminerons les coefficients $\lambda_{i\nu}$ de telle façon que les expressions

$$\frac{D_i^2}{S_i} = \frac{\left(\sum_{\nu=1}^{k} \lambda_{\nu i} d_\nu\right)^2}{\sum_{\nu=1}^{k} \lambda_{\nu i}^2 \sigma_{\nu i}^2} , \quad i = 1, 2 ,$$

atteignent leurs maxima. Par une voie analogue on obtient

$$\lambda_{\nu_i} = \frac{d_\nu}{\sigma_{\nu_i}^2} , \quad \nu = 1, 2, \dots, k ; \quad i = 1, 2 ,$$

et les expressions

(3.1) $$\frac{\sum_{\nu=1}^{k} (\xi_\nu - m_\xi^\nu) \lambda_{\nu_1}}{\sqrt{\sum_{\nu=1}^{k} \lambda_{\nu_1}^2 \sigma_{\nu_1}^2}}$$

(3.2) $$\frac{\sum_{\nu=1}^{k} (\xi_\nu - m_\eta^\nu) \lambda_{\nu_2}}{\sqrt{\sum_{\nu=1}^{k} \lambda_{\nu_2}^2 \sigma_{\nu_2}^2}}$$

suivront la loi de Laplace N (0,1).

Pour l'élément observé à mesures transformées $\xi_1^\circ, \dots, \xi_k^\circ$ qu'on doit classer, c'est-à-dire déterminer s'il appartient à la population π_1 ou à la population π_2, nous obtenons en remplaçant ξ_ν par ξ_ν° dans les ex-

pressions (3.1) et (3.2) deux valeurs numériques λ_{p_1} et λ_{p_2} qui représentent les écarts normaux correspondant aux probabilités p_i %, (i = 1,2). Si λ_{p_i} dépasse le seuil de signification, on rejetera l'hypothèse que l'élément considéré appartient à la population π_i. Les plans

$$\sum_{\nu=1}^{k} (\xi_\nu - m_\xi^\nu) \lambda_{\nu_1} = \lambda_{p_1} \sqrt{S_1}$$

et

$$\sum_{\nu=1}^{k} (\xi_\nu - m_\eta^\nu) \lambda_{\nu_2} = \lambda_{p_2} \sqrt{S_2}$$

ne sont plus parallèles. Si nous prenons de nouveau pour le plan de discrimination le lieu géométrique des points pour lesquels $\lambda_{p_1} = -\lambda_{p_2}$, c'est-à-dire

$$\text{Pr.} (\xi \in \pi_1) = \text{Pr.} (\xi \in \pi_2),$$

son équation sera

$$\sum_{\nu=1}^{k} \left(\frac{\lambda_{\nu_1}}{\sqrt{S_1}} + \frac{\lambda_{\nu_2}}{\sqrt{S_2}} \right) \xi_\nu = \sum_{\nu=1}^{k} \left(\frac{m_\xi^\nu \lambda_{\nu_1}}{\sqrt{S_1}} + \frac{m_\eta^\nu \lambda_{\nu_2}}{\sqrt{S_2}} \right)$$

La différence entre ce plan de discrimination et le précédent (2.9) consiste en ce qui λ_{p_1} $(= -\lambda_{p_2})$ reste constant lorsque ξ se déplace dans le premier, tandis qu'il est variable lorsqu'il se déplace dans le deuxième.

4. DISCRIMINATION DES POPULATIONS DANS LE CAS GÉNÉRAL

Soient π_1 et π_2 deux populations qui s'étendent dans un espace à k dimensions R^k avec les lois de probabilité $f_1 (x_1, \dots, x_k)$ et $f_2 (x_1, \dots, x_k)$ dont la forme peut être quelconque.

Supposons d'abord que nous connaissons a priori les probabilités p_i $(p_1 + p_2 = 1)$ selon lesquelles un élément observé ξ $(\xi_1, \dots, \xi_k)$ appartiendrait à la population π_i. Ce serait, par exemple, la déterminations du sexe des squelettes humains (E.S. Martin (17)). Le problème consiste à séparer aussi bien que possible l'espace R^k en deux régions R_1^k et R_2^k qui s'excluent mutuellement et qui nous permettraient d'accepter l'hypothèse que l'élément ξ appartient à la population π_i s'il se trouve dans la région R_i^k. Nous y arriverons en déterminant les régions R_1^k et R_2^k de telle sorte que l'erreur totale de classification

$$\alpha = p_1 \int_{R_2^k} f_1 \, dV + p_2 \int_{R_1^k} f_2 \, dV$$

soit la plus petite possible. En partant de la probabilité a postériori, selon laquelle l'élément ξ appartenait à la population π_i, c'est-à-dire.

$$\text{Pr.} (\xi \in \pi_i) = \frac{p_i f_i (\xi_1, \dots, \xi_k)}{p_1 f_1 (\xi_1, \dots, \xi_k) + p_2 f_2 (\xi_1, \dots, \xi_k)},$$

on obtient directement

$$p_1 f_1 (\xi_1, \ldots, \xi_k) = p_2 f_2 (\xi_1, \ldots, \xi_k)$$

comme surface de discrimination. Par conséquent, la région R_i^k est définie par l'inégalité $p_i f_i \geq p_j f_j$ $(i \neq j, \quad i, j = 1, 2)$ (Welch (28)).

Si les probabilités p_1 et p_2 sont inconnues, Welch (28), en supposant que la masse de la première population dans R_2^k est égale à la masse de la seconde population dans R_1^k, c'est-à-dire.

$$\int_{R_2^k} f_1 \, dV = \int_{R_1^k} f_2 \, dV ,$$

ou bien

$$\int_{R_1^k} (f_1 + f_2) \, dV = 1 , \tag{4.1}$$

il détermine la région R_1^k de telle façon que l'expression

$$\int_{R_1^k} f_1 \, dV \tag{4.2}$$

atteigne son maximum. Ces hypothèses faites il montre que l'hypersurface séparant les régions R_1^k et R_2^k est donnée par

$$f_1 (x_1, \ldots, x_k) = C f_2 (x_1, \ldots, x_k), \tag{4.3}$$

où la constante C doit être choisie de telle sorte que la condition (4.1) soit satisfaite. C'est cette hypersurface que Welch prend comme la surface de discrimination des populations π_1 et π_2.

En supposant la continuité des lois de probabilité f_1 et f_2 en tout point de R^k, on obtient ce résultat en cherchant le maximum de la fonctionnelle

$$F(R_1^k) = \int_{R_1^k} \left\{ f_1 - \lambda (f_1 + f_2) \right\} dV , \tag{4.4}$$

où le paramètre λ doit être tel que la condition (4.1) soit satisfaite. Soit $P(x_1, \ldots, x_k)$ un point frontière de R_1^k et choisissons arbitrairement un élément ΔR_1^k de R^k de volume $\Delta V > 0$, dont P est un point intérieur. En supposant, par exemple, que ΔR_1^k est situé à l'intérieur d'une sphère centrée en P et de rayon ρ et en faisant tendre ρ, c'est-à-dire ΔV vers zéro, le quotient

$$\frac{F(R_1^k + \Delta R_1^k) - F(R_1^k)}{\Delta V}$$

tendra vers une limite déterminée dont la valeur est

$$\lim_{\Delta V \to 0} \frac{F(R_1^k + \Delta R_1^k) - F(R_1^k)}{\Delta V} =$$

$$= \lim_{\Delta V \to 0} \frac{1}{\Delta V} \int_{\Delta V} \{ f_1 - \lambda (f_1 + f_2) \} dV = f_1 - \lambda (f_1 + f_2) .$$

Or pour que (4.4) soit un maximum, cette limite doit être égale à zéro d'où il résulte que dans ce cas tout point frontière doit satisfaire l'équation

$$f_1 - \lambda (f_1 + f_2) = 0 .$$

En posant $C = \lambda / (1 - \lambda)$, cette équation se ramène à (4.3).

Lorsque les deux populations sont normales et leurs matrices de dispersion égales, l'équation (4.3), qui peut être écrite sous la forme

$$\log. f_1 - \log. f_2 = C_1 \quad ,$$

se réduit à

$$\sum_{\zeta} \sum_{\nu} W^{\zeta\nu} d_{\zeta} x_{\nu} = C_2 \ ,$$

c'est-à-dire au plan de discrimination de Fisher.

Dans le cas où les matrices de dispersion de deux populations normales ne sont pas égales, Cedric A. B. Smith (29) montre que la forme analytique de la surface de discrimination est une forme quadratique.

La valeur minimum de la probabilité

$$\alpha = \int_{R_2^k} f_1 \, dV = \int_{R_1^k} f_2 \, dV \ ,$$

calculée à la manière décrite ci-dessus, représente le degré de confiance relatif à l'hypothèse que l'élément $\xi \in R_i^k$ appartient à la population π_i . Rao (27) élargit cette théorie en examinant le cas lorsque α est au-dessus du seuil de signification. Dans ce cas-là, nous ne pouvons pas accepter l'hypothèse que l'élément ξ de R_i^k appartient à la population π_i , malgré que la surface de discrimination était placée aussi bien que possible. Pour cette raison Rao divise l'espace R^k en quatre régions $R_1^k - C_1$, C_1 , C_2 et $R_2^k - C_2$. En désignant par α_1 le seuil de signification, il est toujours possible de déterminer la région C_1 de telle sorte que

$$\int_{C_1} f_1 \, dV$$

atteigne son maximum sous la condition

$$\int_{C_1} f_2 \, dV = \alpha_1 \ .$$

En appliquant le calcul de variations Rao a obtenu l'inégalité $f_1 \geq Kf_2$ qui définit la région C_1, la constante K étant déterminée de telle sorte que la condition (4.5) soit satisfaite. Si l'élément ξ se trouve dans la région $R_1^k - C_1$, on accepte l'hypothèse qu'il appartient à la population π_1. S'il se trouve dans la région C_1, cette hypothèse ne saurait pas être acceptée. Tout ce que nous pourrions dire à ce sujet-là, c'est qu'il est plus probable qu'il appartient à la population π_1 qu'à la population π_2. D'une façon analogue sera formée la région C_2. Ces deux régions C_1 et C_2 constituent ensemble une région dite "douteuse", car le problème reste sans solution lorsque l'élément examiné se trouve dans cette région. Dans ce cas on mesure les nouveaux caractères, afin d'éviter cette région douteuse.

Remarquons encore que Rao (27, 34, 39) a élargi cette théorie au cas de trois ou plusieurs populations.

5. PLAN GÉNÉRAL DE DISCRIMINATION

La surface de discrimination (4.3) peut avoir une forme bien compliquée. Cependant, il arrive que la partie de cette face, qui partage le gros des masses de π_1 et π_2 ne diffère pas beaucoup d'un plan. Il serait plus pratique dans ce cas-là de former une nouvelle surface de discrimination qui serait toujours un plan.

Soit

$$x_k = \lambda_0 + \sum_{\nu=1}^{k-1} \lambda_\nu x_\nu \tag{5.1}$$

l'équation de ce plan et déterminons les coefficients λ_ν $(\nu = 0, 1, 2, \ldots, k-1)$ de telle sorte que, à la condition que

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \int_{-\infty}^{x_k} (f_1 + f_2)\, dV = 1, \tag{5.2}$$

l'expression

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \int_{-\infty}^{x_k} f_1\, dV \tag{5.3}$$

atteigne son maximum. Désignons par

$$\varphi(x_1, \ldots, x_{k-1}) \equiv \lambda_0 + \lambda_1 x_1 + \ldots + \lambda_{k-1} x_{k-1},$$

$$f_{i1} = f_i(x_1, \ldots, x_{k-1}, \varphi), \quad (i = 1, 2),$$

$$dV_1 = dx_1 \ldots dx_{k-1}$$

et

$$F_i = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} \int_{-\infty}^{x_k} f_i \, dV , \quad (i = 1, 2)$$

de sorte que

$$\frac{\partial F_i}{\partial \lambda_\nu} = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} \frac{\partial \varphi}{\partial \lambda_\nu} f_{i1} \, dV_1 , \quad (i = 1, 2 ; \ \nu = 0, 1, \ldots, k-1).$$

Le système de $(k + 1)$ équations

$$\frac{\partial F_1}{\partial \lambda_\nu} = \lambda \frac{\partial F_2}{\partial \lambda_\nu} , \qquad \nu = 1, 2, \ldots, k ,$$

$$F_1 + F_2 = 1 ,$$

déterminera alors les coefficients cherchés λ_ν $(\nu = 0, 1, \ldots, k - 1)$ ainsi que le paramètre λ .

APPLICATION AUX POPULATIONS NORMALES A DEUX DIMENSIONS. - Soient

(5.5) $$f_1(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left\{\frac{(x-m_1)^2}{\sigma_1^2} - \frac{2\rho(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2}\right\}}$$

et

(5.6) $$f_2(x,y) = \frac{1}{2\pi\sigma_1'\sigma_2'\sqrt{1-\rho'^2}} e^{-\frac{1}{2(1-\rho'^2)}\left\{\frac{(x-m_1')^2}{\sigma_1'^2} - \frac{2\rho'(x-m_1')(y-m_2')}{\sigma_1'\sigma_2'} + \frac{(y-m_2')^2}{\sigma_2'^2}\right\}}$$

les lois de probabilité des populations π_1 et π_2 et

(5.7) $$y = \alpha x + \beta$$

l'équation de la droite de discrimination que l'on cherche. Déterminons les coefficients α et β de telle sorte que la droite (5.7) sépare les populations π_1 et π_2 aussi bien que possible. Le système (5.4) se réduit au système de trois équations suivantes

(5.8) $$\int_{-\infty}^{+\infty} x f_1(x, \alpha x + \beta) \, dx = \lambda \int_{-\infty}^{+\infty} x f_2(x, \alpha x + \beta) \, dx ,$$

$$(5.9)\qquad \int_{-\infty}^{+\infty} f_1(x, \alpha x+\beta)\, dx = \lambda \int_{-\infty}^{+\infty} f_2(x, \alpha x+\beta)\, dx ,$$

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{\alpha x+\beta} \left\{ f_1(x,y) + f_2(x,y) \right\} dx\, dy = 1 .$$

En remplaçant (5.5) et (5.6) dans la troisième équation et en transformant les coordonnées nous obtenons

$$\int_{-\infty}^{+\infty}\int_{\eta_1(\xi)}^{+\infty} \frac{1}{2\pi} e^{-\frac{1}{2}(\xi^2+\eta^2)}\, d\xi\, d\eta = \int_{-\infty}^{+\infty}\int_{-\infty}^{\eta_2(\xi)} \frac{1}{2\pi} e^{-\frac{1}{2}(\xi^2+\eta^2)}\, d\xi\, d\eta ,$$

où

$$(5.10)\qquad \eta_1(\xi) = \sqrt{\frac{1-\rho}{1+\rho}}\, \frac{\alpha\sigma_1+\sigma_2}{\sigma_2-\alpha\sigma_1}\, \xi + \sqrt{\frac{2}{1+\rho}}\, \frac{\alpha m_1 - m_2+\beta}{\sigma_2-\alpha\sigma_1} ,$$

et

$$(5.11)\qquad \eta_2(\xi) = \sqrt{\frac{1-\rho'}{1+\rho'}}\, \frac{\alpha\sigma_1'+\sigma_2'}{\sigma_2'-\alpha\sigma_1'}\, \xi + \sqrt{\frac{2}{1+\rho'}}\, \frac{\alpha m_1' - m_2'+\beta}{\sigma_2'-\alpha\sigma_1'} .$$

De cette façon les fonctions, figurant sous les signes d'intégrales, sont identiques et symétriques par rapport au centre des coordonnées. L'égalité en question sera obtenue si les droites

$$\eta_1(\xi) - \eta = 0$$
$$\eta_2(\xi) - \eta = 0$$

se trouvent à la même distance du centre des coordonnées, c'est-à-dire si

$$(5.12)\qquad \frac{\alpha m_1 - m_2+\beta}{\sqrt{\alpha^2\sigma_1^2 - 2\alpha\rho\sigma_1\sigma_2+\sigma_2^2}} = - \frac{\alpha m_1' - m_2'+\beta}{\sqrt{\alpha^2\sigma_1'^2 - 2\alpha\rho'\sigma_1'\sigma_2'+\sigma_2'^2}}$$

En divisant l'équation (5.8)par l'équation (5.9) on élimine λ et on aura

$$(5.13)\qquad \frac{\int_{-\infty}^{+\infty} x f_1(x, \alpha x+\beta)\, dx}{\int_{-\infty}^{+\infty} f_1(x, \alpha x+\beta)\, dx} = \frac{\int_{-\infty}^{+\infty} x f_2(x, \alpha x+\beta)\, dx}{\int_{-\infty}^{+\infty} f_2(x, \alpha x+\beta)\, dx} .$$

En remplaçant (5.5) et (5.6) dans (5 . 13) et en intégrant nous obtiendrons la deuxième équation par α et β

$$(5.14) \qquad \frac{m_1\sigma_2^2 - \rho\sigma_1\sigma_2(m_1\alpha - m_2+\beta) + (m_2-\beta)\sigma_1^2\alpha}{\sigma_2^2 - 2\rho\sigma_1\sigma_2\alpha+\sigma_1^2\alpha^2} = \\ = \frac{m_1'\sigma_2'^2 - \rho'\sigma_1'\sigma_2'(m_1'\alpha - m_2'+\beta)+(m_2'-\beta)\sigma_1'^2\alpha}{\sigma_2'^2 - 2\rho'\sigma_1'\sigma_2'\alpha+\sigma_1'^2\alpha^2} .$$

Si les matrices de dispersion de π_1 et π_2 sont identiques, c'est-à-dire si $\sigma_1=\sigma_1'$, $\rho=\rho'$ et $\sigma_2=\sigma_2'$, l'équation (5.12) se réduira alors à

$$\alpha m_1 - m_2 + \beta = -\alpha m_1' + m_2' - \beta$$

et l'équation (5.13) à

$$(m_1 - m_1')\sigma_2^2 - \rho\sigma_1\sigma_2\left[(m_1-m_1')\alpha + (m_2-m_2')\right] + (m_2-m_2')\sigma_1^2\alpha = 0$$

de sorte que

$$\alpha = \frac{\rho\sigma_1\sigma_2 d_2 - \sigma_2^2 d_1}{\sigma_1^2 d_2 - \rho\sigma_1\sigma_2 d_1} = \frac{W_{12} d_2 - W_{22} d_1}{W_{11} d_2 - W_{12} d_1}$$

et

$$\beta = \frac{W_{11}(m_2^2 - m_2'^2) - 2W_{12}(m_1 m_2 - m_1' m_2') + W_{22}(m_1^2 - m_1'^2)}{W_{11} d_2 - W_{12} d_1} .$$

On pourrait facilement constater que la droite de discrimination ainsi obtenue coincide avec la droite de Fisher.

CHAPITRE II

6. CLASSIFICATION DES ÉLÉMENTS TRANSFORMÉS

Nous avons vu que la construction de la fonction de discrimination de Fisher était basée sur la différence des moyennes arithmétiques des caractères correspondants de deux populations, de sorte qu'il y aura un contraste entre ces populations à moins que toutes les différences se réduisent à zéro. Cependant, cette méthode est valable lorsqu'on prend sous les conditions identiques les mesures des éléments des populations π_1 et π_2 ainsi que de l'élément examiné ξ, ce qui n'est pas fréquent dans la pratique. C'est pourquoi la méthode de Fisher, qui a été utilisée d'une façon systématique dans bien des problèmes statistiques touchant les domaines les plus différents, ne correspond quelquefois pas à la nature du problème de la discrimination.

Dans ce chapitre nous allons élargir la théorie de discrimination en considérant quelques cas, où les mesures des caractères des éléments $x \in \pi_1$, $x \in \pi_2$ et $x = \xi$ ne sont pas prises dans les mêmes conditions. Lorsque, par exemple, l'élément ξ, avant d'être mesuré, s'est transformé de telle façon que les valeurs de ses caractères ont changé proportionnellement, il est évident que la méthode de Fisher cesse d'être valable. Pour la discrimination des populations π_1 et π_2, ne sont plus essentielles les différences des moyennes arithmétiques mais leurs proportions. Ces considérations nous ont conduit à construire dans le paragraphe 7 une fonction, dite "fonction de disproportion" qui s'adapterait mieux à la nature de ce problème.

On trouve souvent dans la pratique de pareils problèmes. Ainsi, par exemple, les explorateurs scientifiques découvrent souvent des empreintes incrustées dans la pierre ou dans la glace appartenant aux diverses espèces humaines et animales. Etant donné que ces empreintes sont sujettes aux changements du volume, la méthode de la fonction de disproportion y trouverait son application.

Comme un autre exemple supposons deux populations d'adultes π_1 et π_2, telles que les dimensions des crânes de l'espèce π_2 soient visiblement plus petites que celles de l'espèce π_1. Prenons maintenant le cas d'un adolescent appartenant à l'espèce de la population π_1, dont le crâne est approximativement semblable au crâne des adultes de la même espèce. En appliquant la méthode de Fisher, nous risquons d'obtenir une conclusion inexacte, à savoir que le crâne examiné appartient à l'espèce π_2. Par contre, la méthode de la fonction de disproportion tient compte de la ressemblance du crâne de l'adolescent à celui des adultes, et, par conséquent, elle serait plus efficace.

Lorsque les valeurs des caractères de tous les éléments changent proportionnellement, d'après une loi connue de transformation, nous mon-

trerons dans le paragraphe 8, comment il est possible de résoudre le problème de la discrimination au moyen de la méthode de Welch.

Enfin, remarquons que l'exemple de Rao dans le paragraphe 2 était bien fait, car il n'y a pas de modification dans les dimensions des os au cours de la dessication ou de la fossilisation. Cela s'explique par le fait que la substance minérale osseuse forme un réseau tridimensionnel submicroscopique rigide et lorsque disparaissent l'eau et la matière organique contenues dans les mailles de ce réseau, les espaces laissés libres ne modifient ni leur forme ni leurs dimensions, comme le prouvent les études optiques au microscope polarisant (C.A. Baud et M.J. Dallemagne (36)).

7. FONCTION DE DISPROPORTION

Supposons que les valeurs des caractères de l'élément examiné changent proportionnellement de sorte que les mesures $x_1, \ldots, x_k$ se transforment en $\tau x_1, \ldots, \tau x_k$. Afin de classer l'élément transformé par rapport à deux populations connues π_1 et π_2, nous supposerons que ces deux populations sont normales non-singulières et que leurs axes principaux sont parallèles. Soient m_x^i et m_y^i les moyennes arithmétiques du i-ème caractère des populations π_1 et π_2,

$$(w_{x_\zeta x_\nu}) \, , \qquad \zeta, \nu = 1, \ldots, k \, ,$$

$$(w_{y_\zeta y_\nu}) \, , \qquad \zeta, \nu = 1, \ldots, k \, ,$$

leurs matrices de dispersion et

$$f_1(x_1, \ldots, x_k) = \frac{1}{(2\pi)^{k/2}\sqrt{w_x}} \, e^{-\frac{1}{2}\sum_\zeta \sum_\nu w^{x_\zeta x_\nu}(x_\zeta - m_x^\zeta)(x_\nu - m_x^\nu)}$$

$$f_2(x_1, \ldots, x_k) = \frac{1}{(2\pi)^{k/2}\sqrt{w_y}} \, e^{-\frac{1}{2}\sum_\zeta \sum_\nu w^{y_\zeta y_\nu}(x_\zeta - m_y^\zeta)(x_\nu - m_y^\nu)}$$

les lois de probabilité de ces deux populations, où w_x et w_y représentent des valeurs des déterminants de leurs matrices de dispersion et $w^{x_\zeta x_\nu}$ et $w^{y_\zeta y_\nu}$ les éléments des matrices réciproques. Vu la supposition selon laquelle les populations sont normales non-singulières, il résulte qu'il n'existe aucun hyperplan dans l'espace R^k pouvant contenir la masse totale de la population π_i $(i = 1,2)$.

Considérons la fonction linéaire par rapport aux x_ν $(\nu = 1, \ldots, k)$

$$(7.1) \qquad z = \sum_{\nu=1}^{k} \lambda_\nu \frac{x_\nu}{m_x^\nu} \, ,$$

où les coefficients λ_ν satisfont la condition

$$(7.2) \qquad \sum_{\nu=1}^{k} \lambda_\nu = 0 \, .$$

Si $x \in \pi_1$, la moyenne arithmétique est nulle, tandis que la variance devient

$$(7.3) \qquad \mathrm{Var.}_1(z) = \sum_{\zeta=1}^{k} \sum_{\nu=1}^{k} \lambda_\zeta \lambda_\nu C_{\zeta\nu} ,$$

ou, tenant compte de (7.2)

$$(7.4) \quad \mathrm{Var.}_1(z) = \sum_{\zeta=1}^{k-1} \sum_{\nu=1}^{k-1} \lambda_\zeta \lambda_\nu C_{\zeta\nu} - 2\Big(\sum_{\zeta=1}^{k-1} \lambda_\zeta\Big) \sum_{\zeta=1}^{k-1} \lambda_\zeta C_{\zeta k} + \Big(\sum_{\zeta=1}^{k-1} \lambda_\zeta\Big)^2 C_{kk} .$$

Si x coincide avec la moyenne arithmétique de la deuxième population, nous obtenons

$$(7.5) \qquad Q = \sum_{\nu=1}^{k} \lambda_\nu \frac{m_y^\nu}{m_x^\nu} = \sum_{\nu=1}^{k-1} \lambda_\nu (z_\nu - z_k) ,$$

où l'on a posé $z_\nu = m_y^\nu / m_x^\nu$.

Formons le quotient

$$(7.6) \qquad \frac{Q^2}{\mathrm{Var.}_1(z)} .$$

Plus la disproportion des moyennes arithmétiques des populations π_1 et π_2 est grande, plus grande sera la valeur de cette expression. Par conséquent, cette expression peut être considérée comme une mesure de disproportion des moyennes arithmétiques des populations π_1 et π_2. Dans le but de séparer, d'une façon aussi bien que possible, les deux populations, nous déterminerons les coefficients λ_ν de sorte que (7.6) atteigne son maximum. En annulant les dérivées partielles de (7.6) par rapport aux $\lambda_\nu (\nu = 1, \ldots, k-1)$, on obtient le système d'équations.

$$\sum_{\zeta=1}^{k-1} \lambda_\zeta (C_{\zeta\nu} - C_{\zeta k} - C_{k\nu} + C_{kk}) = \frac{\mathrm{Var.}_1(z)}{Q} (z_\nu - z_k) ,$$

$$\nu = 1, \ldots, k-1 .$$

Etant donné que l'expression $\mathrm{Var.}_1(Z)/Q$ figure en facteur, les coefficients cherchés seront proportionnels aux solutions du système linéaire

$$(7.7) \qquad \begin{aligned} R_{11}\lambda_1 + R_{12}\lambda_2 + \ldots + R_{1,k-1}\lambda_{k-1} &= z_1 - z_k \\ R_{21}\lambda_1 + R_{22}\lambda_2 + \ldots + R_{2,k-1}\lambda_{k-1} &= z_2 - z_k \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ R_{k-1,1}\lambda_1 + R_{k-1,2}\lambda_2 + \ldots + R_{k-1,k-1}\lambda_{k-1} &= z_{k-1} - z_k , \end{aligned}$$

où

$$R_{\zeta\nu} = C_{\zeta\nu} - C_{\zeta k} - C_{k\nu} + C_{kk} .$$

Etant donné que $C_{\zeta\nu} = C_{\nu\zeta}$, il sera aussi $R_{\zeta\nu} = R_{\nu\zeta}$. La solution du système (7,7) sera

$$(7.8) \qquad \begin{aligned} \lambda_\zeta &= \sum_{\nu=1}^{k-1} R^{\zeta\nu} (z_\nu - z_k) , \quad \zeta = 1, \ldots, k-1 , \\ \lambda_k &= \sum_{\zeta=1}^{k-1} \sum_{\nu=1}^{k-1} R^{\zeta\nu} (z_k - z_\nu) . \end{aligned}$$

En utilisant (7.8) dans la formule (7.5), on obtient

$$(7{,}9) \qquad Q = \sum_{\zeta=1}^{k-1} \sum_{\nu=1}^{k-1} R^{\zeta\nu} (z_\zeta - z_k)(z_\nu - z_k) ,$$

tandis qu'en multipliant la ν -ème équation du système (7.7) par λ_ν et en additionnant les équations, on aura

$$\sum_{\nu=1}^{k-1} \lambda_\nu \sum_{\zeta=1}^{k-1} \lambda_\zeta R_{\zeta\nu} = \sum_{\nu=1}^{k-1} \lambda_\nu (z_\nu - z_k) ,$$

ou bien

$$\sum_{\zeta=1}^{k-1} \sum_{\nu=1}^{k-1} \lambda_\zeta \lambda_\nu C_{\zeta\nu} - 2 \left(\sum_{\zeta=1}^{k-1} \lambda_\zeta\right) \sum_{\zeta=1}^{k-1} C_{\zeta k} \lambda_\zeta + \left(\sum_{\zeta=1}^{k-1} \lambda_\zeta\right)^2 C_{kk} = \sum_{\nu=1}^{k-1} \lambda_\nu (Z_\nu - Z_k).$$

En utilisant les formules (7.4) et (7.5) il s'ensuit que

$$(7.10) \qquad \mathrm{Var.}_1 (z) = Q ,$$

d'où

$$(7.11) \qquad \mathrm{Max.} \left\{ Q^2 / \mathrm{Var.}_1 (z) \right\} = \sum_{\zeta=1}^{k-1} \sum_{\nu=1}^{k-1} R^{\zeta\nu} (z_\zeta - z_k)(z_\nu - z_k) .$$

L'expression (7.11) représente une mesure de disproportion, c'est-à-dire un contraste statistique entre les populations π_1 et π_2. Cette mesure sera égale à zéro si les moyennes arithmétiques correspondantes sont proportionnelles. L'expression (7.11) joue le même rôle que la distance généralisée D_k de Mahalanobis et, par conséquent, elle peut être utilisée dans la théorie de la séparation des populations en groupes de constellations.

Etant donné que les axes principaux de π_1 et π_2 sont parallèles et que les variables $x_1, \ldots, x_k$ suivent les lois de Laplace, on peut obtenir les variables nouvelles $\xi_1, \ldots, \xi_k$ qui sont indépendantes en suivant également les lois de Laplace. En appliquant le même procédé à ces nouvelles variables, nous obtiendrons les coefficients λ_ν, tandis que l'expression

$$(7.12) \qquad \frac{\sum_{\nu=1}^{k} \lambda_\nu \dfrac{\xi_\nu - m_\xi^\nu}{m_\xi^\nu}}{\sqrt{\sum_{\nu=1}^{k} \lambda_\nu^2 \dfrac{\sigma_{\xi\nu}^2}{m_\xi^{\nu 2}}}} = \frac{z_1}{\sigma_{z_1}} .$$

où m_ξ^ν et σ_ξ^ν sont les paramètres transformés de la population π_1 aura la distribution normale N (0,1).

En introduisant le rapport réciproque m_ξ^ν / m_η^ν dans la fonction de disproportion, on obtient par le même procédé les coefficients correspondants λ'_ν, tandis que l'expression

$$(7.13) \qquad \frac{\sum_{\nu=1}^{k} \lambda'_\nu \dfrac{\xi_\nu - m_\eta^\nu}{m_\eta^\nu}}{\sqrt{\sum_{\nu=1}^{k} \lambda'^2_\nu \dfrac{\sigma_{\eta\nu}^2}{m_\eta^{\nu 2}}}} = \frac{z_2}{\sigma_{z_2}} .$$

où m_ξ^ν et σ_{η_ν} sont les paramètres transformés de la population π_2, suivra également la loi normale N (0,1).

Si nous avons à présent un élément avec des mesures transformées $\xi_1^0, \ldots, \xi_k^0$ qu'on veut classer en sachant qu'il ne peut appartenir qu'à

l'une des populations π_1 et π_2 et que cet élément a subi des changements proportionnels de ses caractères, avant d'être mesuré, nous obtiendrons - en remplaçant ξ_ν par ξ°_ν dans les formules (7 . 12) et (7 . 13) - deux valeurs numériques λ_{p_1} et λ_{p_2} qui représentent les écarts normaux correspondants aux probabilités de p_i % (i = 1,2). Si λ_{p_i} dépasse le seuil de signification, on rejettera l'hypothèse que l'élément considéré appartient à la population π_i . En prenant pour le plan de discrimination le lieu géométrique des points pour lesquels $\lambda_{p_1} = -\lambda_{p_2}$, c'est-à-dire

$$\Pr.(\xi \in \pi_1) = \Pr.(\xi \in \pi_2)$$

son équation sera

$$z_1 \sigma_{z_2} + z_2 \sigma_{z_1} = 0 .$$

8. CLASSIFICATION D'UN ÉLÉMENT TRANSFORMÉ PAR RAPPORT A DEUX POPULATIONS TRANSFORMÉES

Soit $f_1(\xi_1, \dots, \xi_k)$ la loi de probabilité de la population π_1 et supposons que, sous l'influence de certaines causes, les caractères de tous les éléments ont changé proportionnellement, c'est-à-dire que

$$\xi_\nu = \tau x_\nu \quad , \qquad \nu = 1, \dots, k ,$$

où le facteur de proportionnalité τ est une variable aléatoire qui peut varier dans le cas général de 0 à ∞ . Soit $\rho(\tau)$ sa loi de probabilité et $\psi(t)$ sa fonction caractéristique. On obtient la population transformée π_2 en ajoutant aux éléments de la population primitive π_1 tous les éléments qui leur sont semblables, tout en tenant compte de la loi de probabilité du facteur de proportionnalité. Désignons par $f_i(x) = f_i(x_1 \dots x_k)$ la loi de probabilité de la population π_i , par $F_i(x) = F_i(x_1 \dots x_k)$ sa fonction de répartition et par $\varphi_i(t) = \varphi_i(t_1 \dots t_k)$ sa fonction caractéristique.

Soient $(\xi_1, \dots, \xi_k)$ et τ les variables indépendantes dont les moyennes arithmétiques sont finies et supposons que l'intégrale

$$\int_1^\infty \left| \frac{\psi(t)}{t} \right| dt$$

converge. En utilisant la formule de H. Cramer (35) pour la détermination de la loi de répartition du quotient de deux variables aléatoires, nous pouvons introduire la fonction de répartition du quotient à k dimensions $\frac{\xi}{\tau} \left\{ \frac{\xi_1}{\tau}, \dots, \frac{\xi_k}{\tau} \right\}$

$$F_2(x) = \left(-\frac{1}{2\pi i}\right)^k \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \frac{\varphi_1(t)\,\psi(-tx) - \psi(t_1)\dots\psi(t_k)}{t_1 \dots t_k}\, dt_1 \dots dt_k ,$$

où l'on a posé $tx = t_1 x_1 + \dots + t_k x_k$. En supposant que nous pouvons dériver cette expression par rapport aux variables $x_1, \dots, x_k$, on obtient

$$f_2(x) = \left(-\frac{1}{2\pi i}\right)^k \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \frac{\varphi_1(t)\,\frac{\partial}{\partial x}\,\psi(-tx)}{t_1 \dots t_k}\, dt_1 \dots dt_k ,$$

où $\frac{\partial}{\partial x} = \frac{\partial^k}{\partial x_1 \ldots \partial x_k}$. En dérivant ψ (-tx) et en échangeant l'ordre des intégrations, nous aurons

$$f_2(x) = \frac{1}{(2\pi)^k} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \varphi_1(t) \int_0^{+\infty} e^{-i\tau t x} \tau^k \rho(\tau)\, d\tau\, dt_1 \ldots dt_k$$

$$= \frac{1}{(2\pi)^k} \int_0^{\infty} \tau^k \rho(\tau)\, d\tau \int_{-\infty}^{+\infty} \cdots \int^{+\infty} e^{-i\tau t x} \varphi_1(t)\, dt_1 \ldots dt_k ,$$

ou bien

$$(8.1) \qquad f_2(x) = \int_0^{\infty} f_1(\tau x)\, \tau^k \rho(\tau)\, d\tau ,$$

c'est-à-dire

$$f_2(x_1, \ldots, x_k) = \int_0^{\infty} f_1(\tau x_1, \ldots, \tau x_k) \tau^k \rho(\tau) d\tau .$$

On voit que la loi de probabilité f_2 (x) est une fonctionnelle de f_1 (x) et nous pouvons écrire

$$f_2(x) = \mathcal{F}\left[f_1\right].$$

Cette fonctionnelle est linéaire, car on a

$$\mathcal{F}\left[af_1^1 + bf_1^2\right] = a\,\mathcal{F}\left[f_1^1\right] + b\,\mathcal{F}\left[f_1^2\right]$$

quelles que soient les lois de probabilité f_1^1 (x) et f_1^2 (x) et quels que soient les nombres a et b.

La fonction f_2 (x) représente une loi de probabilité, car la formule (8.1) nous montre quelle est positive et que

$$(8.2) \qquad \int_{R^k} f_2(x)\, dx_1 \ldots dx_k = \int_{R^k} \int_0^{\infty} f_1(\tau x)\, \tau^k \rho(\tau)\, d\tau\, dx_1 \ldots dx_k = 1 .$$

Supposons qu'une nouvelle transformation des caractères, identique à la précédente, se soit produite. La loi de probabilité de la population transformée π_3 sera

$$f_3(x) = \mathcal{F}\left[f_2\right] = \int_0^{+\infty} \int_0^{+\infty} (\tau_1\tau_2)^k \rho(\tau_1)\rho(\tau_2) f_1(\tau_1\tau_2 x)\, d\tau_1\, d\tau_2 .$$

En général, si la population π_1 a subi en tout n transformations succes-

sives de sorte que chaque fois, lorsqu'on passe de π_i à π_{i+1}, les valeurs des caractères changent proportionnellement avec la même loi $\rho(\tau)$, la population π_n, aura comme sa loi de probabilité

$$(8.4)\quad f_{n+1}(x) = \mathcal{F}\left[f_n\right] = \int_0^{+\infty}\cdots\int_0^{+\infty}(\tau_1\ldots\tau_n)^k\rho(\tau_1)\ldots\rho(\tau_n)\,f(\tau_1\ldots\tau_n x)\,d\tau_1\ldots d\tau_n.$$

En remplaçant

$$\tau_i = t_i \quad , \quad i = 1, \ldots, n-1 ,$$
$$\tau_1\ldots\tau_n = t$$

la formule (8.4) devient

$$f_{n+1}(x) = \int_0^{+\infty}\cdots\int_0^{+\infty} t^k\rho(t_1)\ldots\rho(t_{n-1})\,\rho\left(\frac{t}{t_1\ldots t_{n-1}}\right) f_1(tx)\,\frac{dt_1}{t_1}\ldots\frac{dt_{n-1}}{t_{n-1}}\,dt ,$$

ou bien

$$(8.5)\qquad f_{n+1}(x) = \int_0^{+\infty} t^n\rho_n(t)\,f_1(tx)\,dt$$

où l'on a posé

$$\rho_n(t) = \int_0^{+\infty}\cdots\int_0^{+\infty}\frac{\rho(t_1)}{t_1}\cdots\frac{\rho(t_{n-1})}{t_{n-1}}\,\rho\left(\frac{t}{t_1\ldots t_{n-1}}\right)dt_1\ldots dt_{n-1},$$
$$\rho_1(t) = \rho(t) .$$

Etant donné que la fonction $\rho_n(t)$ est positive et que

$$\int_0^{+\infty}\rho_n(t)\,dt = 1 ,$$

il en résulte que $\rho_n(t)$ représente une loi de probabilité, celle du facteur de proportionnalité en passant directement de π_1 à π_n. Par conséquent, les formules (8.1) et (8.5) ayant la même forme, nous pouvons toujours supposer qu'on n'avait qu'une seule transformation. C'est pourquoi, dans la suite de notre exposé, nous resterons dans le cas n = 1.

Exemple. - Soit

$$f_1(x,y) = \frac{a^2 b^2}{a+b}(x+y)\,e^{-(ax+by)} , \qquad a > 0,\ b > 0,\quad x \geq 0,\ y \geq 0,$$

la loi de probabilité de la population π_1 et

$$\rho(\tau) = \frac{\alpha^\nu}{\Gamma(\nu)}\tau^{\nu-1}e^{-\alpha\tau} , \qquad , \alpha > 0 ,\ \tau \geq 0 ,$$

celle du facteur de proportionnalité τ . En appliquant la formule (8.1) nous obtiendrons

$$f_2(x,y) = \frac{a^2 b^2}{a+b}\,\frac{\alpha^\nu}{\Gamma(\nu)}(x+y)\int_0^\infty \tau^{\nu+2}\, e^{-\tau(ax+by+\alpha)}\, d\tau =$$

$$= \frac{a^2 b^2}{a+b}\,\frac{\alpha^\nu}{\Gamma(\nu)}\Gamma(\nu+3)\,\frac{x+y}{(ax+by+\alpha)^{\nu+3}}\ , x>0\ , y>0 .$$

FONCTION CARACTERISTIQUE ET MOMENTS DE LA POPULATION π_2 .- En désignant la fonction caractéristique de π_1 par φ_1 (t), où $t = t(t_1, \ldots, t_k)$ représente également un point de l'espace R^k, on a

$$(8.6) \qquad \varphi_1(t) = \int_{R^k} e^{i(tx)} f_1(x)\, dx_1 \ldots dx_k$$

De même

$$\varphi_2(t) = \int_{R^k} e^{i(tx)} f_2(x)\, dx_1 \ldots dx_k$$

sera la fonction caractéristique de la population transformée π_2 . En utilisant (8.1) cette fonction peut être écrite sous la forme suivante

$$\varphi_2(t) = \int_0^\infty\int_{R^k} e^{i(tx)} f_1(\tau x)\, \tau^k \rho(\tau)\, d\tau\, dx_1 \ldots dx_k =$$

$$= \int_0^\infty \rho(\tau)\, d\tau \int_{R^k} e^{i\frac{1}{\tau}(t\xi)} f_1(\xi)\, d\xi_1 \ldots d\xi_k$$

ou bien

$$(8.7) \quad \varphi_2(t) = \int_0^\infty \varphi_1\left(\frac{t}{\tau}\right) \rho(\tau)\, d\tau .$$

En développant la fonction φ_1 (t) en série aux environs du point $t_1 = \ldots = t_k = 0$ on aura

$$\int_0^\infty \varphi_1\left(\frac{t}{\tau}\right)\rho(\tau)\, d\tau = \int_0^\infty \rho(\tau)\left\{1 + \frac{i}{\tau}(m_1^1 t_1 + \ldots + m_k^1 t_k) + \right.$$

$$\left. + \frac{i^2}{2!\,\tau^2}\sum_j\sum_s m_{js}^1\, t_j t_s + 0\left(\sum_j t_j^2\right)\right\} d\tau \ ,$$

ou bien

$$(8.8) \qquad \varphi_2(t) = 1 + \frac{i}{1!}\, a_{-1}\sum_j m_j^1 t_j + \frac{i^2}{2!}\, a_{-2}\sum_j\sum_s m_{js}^1\, t_j t_s + \ldots$$

où nous avons désigné par m_j et m_{js}, $(j, s = 1, \ldots, k)$, les moments non-centrés du premier et du deuxième ordre de la loi de probabilité f_1 et par a_{-i} les moments négatifs de $\rho(\tau)$, c'est-à-dire

$$a_{-i} = \int_0^\infty \frac{1}{\tau^i} \rho(\tau) \, d\tau , \qquad i = 1, \ldots, k .$$

De (8.8) nous obtenons les moments non-centrés de la loi de probabilité de la population π_2

$$m_j^2 = a_{-1} m_j^1 , \qquad j = 1, \ldots, k ,$$

$$m_{js}^2 = a_{-2} m_{js}^1 , \qquad j, s = 1, \ldots, k .$$

Surface de discrimination. - Supposons que nous avons constitué les lois de probabilité $f_1 (x_1, \ldots, x_k)$ et $f_2(x_1, \ldots, x_k)$ en mesurant sous les conditions A les éléments des populations π_1 et π_2. Si nous avons pris sous les conditions B les mesures d'un élément x° qu'on veut classer, c'est-à-dire examiner s'il appartient à π_1 où π_2, il est évident que nous ne pouvons pas appliquer la méthode de Welch en utilisant directement les lois f_1 et f_2. Lorsque le passage des conditions A aux conditions B provoque un changement proportionnel des caractères de tous les éléments et lorsque nous connaissons les lois de probabilité du facteur de proportionnalité τ, c'est-à-dire $\rho_1(\tau)$ et $\rho_2(\tau)$, les lois de probabilité des deux populations correspondant aux conditions B seront alors $F_1 [f_1]$ et $F_2 [f_2]$ qui nous serviront pour l'application de la méthode de Welch. La surface de discrimination, séparant les régions R_1^k et R_2^k, sera donnée par

$$F_1 \left[f_1 \right] = C \, F_2 \left[f_2 \right] , \tag{8.9}$$

où la constante C doit être choisie de telle façon que la condition

$$\int_{R_1^k} \left(F_1 \left[f_1 \right] + F_2 \left[f_2 \right] \right) \, dV = 1 \tag{8.10}$$

soit satisfaite. Si l'erreur de classification

$$\alpha = \int_{R_1^k} F_2 \left[f_2 \right] \, dV$$

est inférieure ou égale au seuil de signification, nous accepterons l'hypothèse que x° appartient à la population π_i dans le cas où il se trouve dans la région R_i^k.

Lorsque les lois $\rho_1(\tau)$ et $\rho_2(\tau)$ sont identiques, (8.9) et (8.10) deviennent

$$F \left[f_1 \right] = C \, F \left[f_2 \right]$$

et

$$\int_{R_1^k} \left(F \left[f_1 \right] + F \left[f_2 \right] \right) \, dV = 1 ,$$

ou bien, vu la linéarité de l'opérateur intégral F ,

$$F\left[f_1 - Cf_2\right] = 0$$

et

$$\int_{R_1^k} F\left[f_1 + f_2\right] dV = 1 .$$

Enfin considérons une population π dont une partie π_1 reste constante tandis que les caractères de l'autre $\pi_2 = \pi - \pi_1$, avec la même répartition, commencent à changer proportionnellement à partir du moment $t = t_0$. La loi ρ dépendra alors de τ et t. Soit $f_1(x_1, \ldots, x_k)$ la loi de probabilité de π_1 et

$$f_2(x_1, \ldots, x_k, t) = F\left[f_1\right] = \int_0^\infty \tau^k \rho(\tau, t)\, f_1(\tau x_1, \ldots, \tau x_k)\, d\tau, \quad t > t_0,$$

$$f_2(x_1, \ldots, x_k, t_0) = f_1(x_1, \ldots, x_k)$$

celle de π_2. La surface de discrimination

$$f_1 - CF\left[f_1\right] = 0$$

$$\int_{R_1^k} \left(f_1 + F\left[f_1\right] \right) dV = 1 ,$$

se transforme au cours du temps. Si α représente le seuil de signification et si l'équation

$$\int_{R_1^k} f_1 \, dV = 1 - \alpha$$

a une seule racine réelle $t = t_1 > t_0$, t_1 représentera alors le moment à partir duquel l'appartenance de chaque élément examiné pourra être statistiquement déterminée.

CHAPITRE III

9. DÉTERMINATION DES PARAMÈTRES DE POPULATIONS LORSQUE LA SURFACE DE DISCRIMINATION EST CONNUE

Supposons que nous connaissons les formes analytiques des lois de probabilité de π_1 et π_2 contenant les mêmes paramètres inconnus α_i $(i = 1, 2, \ldots, p)$. En faisant varier les paramètres α_i on obtient une infinité de couples de lois de probabilité, chaque couple déterminant une surface de discrimination de Welch.

Si la surface de discrimination est préalablement fixée, la question qui se pose est de savoir comment faut-il déterminer les valeurs de paramètres α_i pour que les distributions soient telles que la surface de discrimination posée les sépare de la meilleure façon possible.

Soit fixée, par exemple, la frontière entre deux Etats et soient A et B deux villes frontières voisines, A appartenant au pays I et B au pays II. Plus le trafic et l'échange des biens entre ces deux États sont intenses, plus le développement de la population de la ville A dépendra du développement de la population de la ville B. C'est pourquoi les densités des populations des villes A et B dépendront des mêmes paramètres. Quant à la forme mathématique de leurs distributions elle peut être construite approximativement en vertu des conditions géographiques, économiques etc., qui règnent dans la région en question. Au cours du développement des villes A et B, il peut arriver qu'une partie de la ville A se trouve dans l'État voisin, c'est-à-dire sur le territoire de la ville B et vice versa. Ce fait peut produire des conséquences défavorables sur les rapports entre les deux pays. Il serait dans l'intéret de ces deux pays d'organiser le développement des villes A et B de telle façon que la frontière existante les sépare toujours aussi bien que possible tout en tenant compte de leur dépendance réciproque. La question qui se pose ici est de savoir ce qu'on doit entendre sous l'expression "la séparation aussi bonne que possible". La réponse logique et juste qu'on pourrait donner à cette question serait la suivante : réduire au minimum le nombre d'habitants de la ville A qui se trouveraient dans le pays II ainsi que le nombre d'habitants de B dans le pays I, avec la condition que ces deux nombres soient égaux entre eux.

Soient dans le cas général $f_1(x_1, \ldots, x_k\,; \alpha_1, \ldots, \alpha_p)$ et $f_2(x_1, \ldots, x_k; \alpha_1, \ldots, \alpha_p)$ les lois de probabilité des populations π_1 et π_2 à k dimensions. Si la surface fixée de discrimination sépare l'espace R^k en deux régions R_1^k et $R_2^k = R^k - R_1^k$, qui s'excluent mutuellement, notre problème se réduit alors à déterminer les valeurs des paramètres α_i $(i = 1, \ldots, p)$, de sorte qu'à la condition

$$(9.1) \qquad \int_{R_1^k} f_2 \, dV = \int_{R_2^k} f_1 \, dV$$

l'expression

$$(9.2)\qquad \int_{R_1^k} f_1 \, dV$$

atteigne son maximum. Si les dérivées partielles $\frac{\partial f_1}{\partial \alpha_i}$ et $\frac{\partial f_2}{\partial \alpha_i}$ existent, et si nous pouvons admettre que la dérivée partielle de l'intégrale par rapport aux paramètres α_i est égale à l'intégrale de la dérivée partielle de la fonction sous le signe de l'intégrale, se système suivant de (p + 1) équations

$$(9.3)\qquad \begin{gathered} \int_{R_1^k} (f_1 + f_2)\, dV = 1 \\ (1-\lambda)\int_{R_1^k} \frac{\partial f_1}{\partial \alpha_i}\, dV - \lambda \int_{R_1^k} \frac{\partial f_2}{\partial \alpha_i}\, dV = 0, \qquad i = 1, \dots, p, \end{gathered}$$

nous permettra de déterminer les valeurs des paramètres inconnus α_i $(i = 1, \dots, p)$ pour lesquelles l'expression (9.2) atteint son maximum.

EXEMPLE A UNE DIMENSION. - Soient

$$(9.4)\qquad f_1(x;\tau,t) = \frac{1}{\sigma_1(t)\sqrt{2\pi}}\, e^{-\frac{\{x-m_1(\tau)\}^2}{2\sigma_1^2(t)}}$$

$$(9.5)\qquad f_2(x;\tau,t) = \frac{1}{\sigma_2(t)\sqrt{2\pi}}\, e^{-\frac{\{x-m_2(\tau)\}^2}{2\sigma_2^2(t)}}$$

les lois de probabilité de π_1 et π_2 et $x = \alpha$ le point de discrimination. En supposant que

$$(9.6)\qquad \int_{-\infty}^{\alpha} f_2(x;\tau,t)\, dx = \int_{\alpha}^{+\infty} f_1(x;\tau,t)\, dx,$$

déterminons les paramètres τ et t de telle sorte que

$$(9.7)\qquad \int_{-\infty}^{\alpha} f_1(x;\tau,t)\, dx$$

atteigne son maximum. En remplaçant (9.4) et (9.5) dans (9.6), il vient

$$\frac{1}{\sigma_2\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-\frac{(x-m_2)^2}{2\sigma_2^2}}\, dx = \frac{1}{\sigma_1\sqrt{2\pi}} \int_{\alpha}^{+\infty} e^{-\frac{(x-m_1)^2}{2\sigma_1^2}}\, dx,$$

ou

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\alpha-m_2}{\sigma_2}} e^{-\frac{\xi^2}{2}}\, d\xi = \frac{1}{\sqrt{2\pi}} \int_{\frac{\alpha-m_1}{\sigma_1}}^{+\infty} e^{-\frac{\xi^2}{2}}\, d\xi$$

Etant donné que la fonction $e^{-\frac{\xi^2}{2}}$ est paire, nous obtenons l'égalité ci-dessus si

(9.8) $$\frac{\alpha - m_1(\tau)}{\sigma_1(t)} = - \frac{\alpha - m_2(\tau)}{\sigma_2(t)}$$

Désignons par

(9.9) $$F = \int_{-\infty}^{\alpha} f_1\, dx = \int_{-\infty}^{\frac{\alpha - m_1}{\sigma_1}} \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}}\, d\xi$$

et

(9.10) $$\Phi = \int_{-\infty}^{\alpha} f_2\, dx = \int_{-\infty}^{\frac{\alpha - m_2}{\sigma_2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}}\, d\xi \ .$$

Le maximum de (9.7) à condition (9.6) représente un extremum lié. C'est pourquoi les valeurs t et τ seront obtenus de l'équation (9.8) et de

$$(1-\lambda)\frac{\partial F}{\partial \tau} + \lambda \frac{\partial \Phi}{\partial \tau} = 0$$

$$(1-\lambda)\frac{\partial F}{\partial t} + \lambda \frac{\partial \Phi}{\partial \tau} = 0$$

ou, en éliminant λ,

(9.11) $$\frac{\partial F / \partial t}{\partial F / \partial \tau} = \frac{\partial \Phi / \partial t}{\partial \Phi / \partial \tau}$$

Etant donné que de (9.9) et (9.10) nous avons

$$\frac{\partial F}{\partial \tau} = - \frac{m_1'}{\sigma_1 \sqrt{2\pi}} e^{-\frac{(\alpha - m_1)^2}{2\sigma_1^2}} ,$$

$$\frac{\partial F}{\partial t} = - \frac{(\alpha - m_1)\sigma_1'}{\sigma_1^2 \sqrt{2\pi}} e^{-\frac{(\alpha - m_1)^2}{2\sigma_1^2}} ,$$

$$\frac{\partial \Phi}{\partial \tau} = - \frac{m_2'}{\sigma_2 \sqrt{2\pi}} e^{-\frac{(\alpha - m_2)^2}{2\sigma_2^2}} ,$$

$$\frac{\partial \Phi}{\partial t} = - \frac{(\alpha - m_2)\sigma_2'}{\sigma_2^2 \sqrt{2\pi}} e^{-\frac{(\alpha - m_2)^2}{2\sigma_2^2}} ,$$

la formule (9.11) deviendra

$$\frac{\alpha - m_1}{\sigma_1} \frac{\sigma_1'}{m_1'} = \frac{\alpha - m_2}{\sigma_2} \frac{\sigma_2'}{m_2'} ,$$

ou, si nous prenons en considération (9.8),

(9.12) $$\frac{\sigma_1'(t)}{m_1'(\tau)} = - \frac{\sigma_2'(t)}{m_2'(\tau)} \ .$$

Le système des équations (9.8) et (9.12) donne les valeurs cherchées pour les paramètres t et τ .

Si nous prenons, par exemple, que

$$m_1 = \alpha - \frac{1}{1+\tau^2}, \qquad \sigma_1 = 1 + t^2 ,$$

$$m_2 = \alpha + \frac{1}{1+2\tau+\tau^2}, \qquad \sigma_2 = 1+t+t^2 ,$$

l'équation (9.12) se réduira à

$$t(2+2\tau)(1+\tau^2)^2 = \tau(1+2t)(1+2\tau+\tau^2)^2$$

et l'équation (9.8) à

$$(1+t^2)(1+\tau^2) = (1+2\tau+\tau^2)(1+t+t^2).$$

Nous voyons immédiatement qu'une des solutions sera $t = \tau = o$. Il est facile de vérifier que l'extrémum est un maximum, dans ce cas, dont la valeur est

$$\text{Max.}\left\{\int_{-\infty}^{\alpha} f_1(x)\,dx\right\} = \Phi(1).$$

Les moyennes arithmétiques et les écarts types se réduisent à

$$m_1 = \alpha - 1, \quad \sigma_1 = 1, \quad m_2 = \alpha + 1 \quad \text{et } \sigma_2 = 1.$$

EXEMPLE A DEUX DIMENSIONS. - Soient

$$(9.13) \qquad f_1(x,y;\tau,t) = \frac{1}{2\pi}\, e^{-\frac{1}{2}\{x-m_1(t,\tau)\}^2 - \frac{1}{2}\{y-m_2(t,\tau)\}^2},$$

$$(9.14) \qquad f_2(x,y;\tau,t) = \frac{1}{2\pi}\, e^{-\frac{1}{2}\{x-n_1(t,\tau)\}^2 - \frac{1}{2}\{y-n_2(t,\tau)\}^2},$$

les lois de probabilité de π_1 et π_2 et $y = ax + b$ la droite de discrimination. En supposant que

$$(9.15) \qquad \int_{-\infty}^{+\infty}\int_{ax+b}^{+\infty} f_1(x,y;\tau,t)\,dx\,dy = \int_{-\infty}^{+\infty}\int_{-\infty}^{ax+b} f_2(x,y;\tau,t)\,dx\,dy$$

déterminons les paramètres t et τ de telle sorte que

$$(9.16) \qquad \int_{-\infty}^{+\infty}\int_{-\infty}^{ax+b} f_1(x,y;\tau,t)\,dx\,dy$$

atteigne son maximum. En remplaçant (9.13) et (9.14) dans (9.15), il vient

$$\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-m_1)^2}\int_{ax+b-m_2}^{+\infty} e^{-\frac{1}{2}\eta^2}\,d\eta\,dx = \int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-n_1)^2}\int_{-\infty}^{ax+b-n_2}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\eta}\,d\eta$$

où

$$(9.17)\quad \int_{-\infty}^{+\infty}\int_{a(\xi+m_1)+b-m_2}^{+\infty} \frac{1}{2\pi} e^{-\frac{\xi^2+\eta^2}{2}}\, d\xi\, d\eta = \int_{-\infty}^{+\infty}\int_{-\infty}^{a(\xi+n_1)+b-n_2} \frac{1}{2\pi} e^{-\frac{\xi^2+\eta^2}{2}}\, d\xi\, d\eta .$$

Etant donné que les lignes de niveau de la fonction figurant sous le signe de l'intégrale sont des cercles centrés tandis que les frontières sont des droites parallèles, l'équation (9.17) sera satisfaite si les segments de ces droites sur l'axe de l'ordonnée sont égaux par leur valeur absolue mais du signe contraire, c'est-à-dire

$$(9.18)\quad am_1(\tau,t) + b - m_2(\tau,t) = -an_1(\tau,t) - b + n_2(\tau,t) .$$

Désignons par

$$F = \int_{-\infty}^{+\infty}\int_{-\infty}^{ax+b} f_1\, dx\, dy = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} \int_{-\infty}^{a(\xi+m_1)+b-m_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta^2}{2}}\, d\xi\, d\eta ,$$

$$\Phi = \int_{-\infty}^{+\infty}\int_{-\infty}^{ax+b} f_2\, dx\, dy = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} \int_{-\infty}^{a(\xi+n_1)+b-n_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{\eta^2}{2}}\, d\eta\, d\xi ,$$

et cherchons leurs dérivées partielles par rapport à t et τ :

$$\frac{\partial F}{\partial t} = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} \left(a\frac{\partial m_1}{\partial t} - \frac{\partial m_2}{\partial t}\right) \frac{1}{\sqrt{2\pi}} e^{-\frac{\{a(\xi+m_1)+b-m_2\}^2}{2}}\, d\xi =$$

$$= \left(a\frac{\partial m_1}{\partial t} - \frac{\partial m_2}{\partial t}\right) \frac{1}{\sqrt{2\pi}} e^{-\frac{Q}{2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{a^2+1}{2}\left\{\xi + \frac{a(am_1+b-m_2)}{a^2+1}\right\}^2} d\xi ,$$

où l'on a posé

$$Q = a^2 m_1^2 + 2am_1(b-m_2) + (b-m_2)^2 - \frac{a^2(am_1+b-m_2)^2}{a^2+1} =$$

$$= (am_1 + b - m_2)^2 / (a^2+1) = d^2$$

La valeur de l'intégrale étant $(a^2+1)^{-\frac{1}{2}}$, nous avons

$$\frac{\partial F}{\partial t} = \left(a\frac{\partial m_1}{\partial t} - \frac{\partial m_2}{\partial t}\right) \frac{1}{\sqrt{2\pi(a^2+1)}} e^{-d^2/2} .$$

D'une façon analogue on obtient

$$\frac{\partial F}{\partial \tau} = \left(a\frac{\partial m_1}{\partial \tau} - \frac{\partial m_2}{\partial \tau}\right) \frac{1}{\sqrt{2\pi(a^2+1)}} e^{-d^2/2} ,$$

$$\frac{\partial \Phi}{\partial t} = \left(a \frac{\partial n_1}{\partial t} - \frac{\partial n_2}{\partial t} \right) \frac{1}{\sqrt{2\pi (a^2+1)}} e^{-d^2/2} ,$$

$$\frac{\partial \Phi}{\partial \tau} = \left(a \frac{\partial n_1}{\partial \tau} - \frac{\partial n_2}{\partial \tau} \right) \frac{1}{\sqrt{2\pi (a^2+1)}} e^{-d^2/2} ,$$

de sorte que l'équation (9.11) se réduit à

$$(9.19) \qquad \frac{a \frac{\partial m_1}{\partial t} - \frac{\partial m_2}{\partial t}}{a \frac{\partial m_1}{\partial \tau} - \frac{\partial m_2}{\partial \tau}} = \frac{a \frac{\partial n_1}{\partial t} - \frac{\partial n_2}{\partial t}}{a \frac{\partial n_1}{\partial \tau} - \frac{\partial n_2}{\partial \tau}} .$$

Le système des équations (9.18) et (9.19) donne des valeurs cherchées des paramètres t et τ.

CHAPITRE IV

10. INTERDÉPENDANCE DE POPULATIONS

Supposons qu'on ait deux ou plusieurs populations ; lorsque nous examinons l'appartenance d'un élément à l'une de ces populations, il est dans la nature du problème que ces populations soient de la même espèce et qu'elles ne diffèrent entre elles que dans l'accent de certains caractères. En d'autres termes, toutes ces populations doivent appartenir à un plus large groupe de classification, sinon la nature elle-même de l'élément observé (même sans mesurer d'une façon précise ses caractères particuliers) serait déjà suffisante pour déterminer l'appartenance de cet élément. Aussi nous pouvons supposer qu'il existe une corrélation entre les populations observées. Dans le cas de deux populations à chaque changement de la première population π_1 correspond alors en général un changement de la seconde population π_2. Une telle dépendance peut s'exprimer par une relation entre les lois de probabilité des populations π_1 et π_2, et nous supposerons, enfin, que la loi de probabilité de la population π_2 s'obtient de la loi de probabilité de la populations π_1 par une transformation linéaire. Nous verrons que ce cas apparait souvent dans les problèmes statistiques. Si l'on mesure un seul caractère et si $f_2(x)$ est la loi de probabilité de la population π_i, $(i = 1,2)$ on aura alors

$$f_2(x) = \int_{-\infty}^{+\infty} f_1(\xi)\, k(x,\xi)\, d\xi. \tag{10.1}$$

Pour que la fonction $f_2(x)$ ainsi définie, soit une loi de probabilité, il faut que le noyau $k(x,\xi)$ satisfasse à certaines conditions et, en particulier, il suffit que $k(x,\xi)$ soit telle que $f_2(x)$, c'est-à-dire que

$$\int_{-\infty}^{+\infty} f_1(\xi)\, k(x,\xi)\, d\xi \tag{10.2}$$

soit intégrable dans toute intervalle fini, et que

$$k(x,\xi) \geqslant 0, \tag{10.3}$$

$$\int_{-\infty}^{+\infty} k(x,\xi)\, dx = 1. \tag{10.4}$$

En effet, de (10.3) il résulte en premier lieu que $f_2(x) \geqslant 0$, étant donné que $f_1(x) \geqslant 0$.

En second lieu, afin de montrer que $f_2(x)$ est intégrable dans l'intervalle $(-\infty, +\infty)$ et que la valeur de cette intégrale est égale à 1, consi-

dérons l'intégrale

$$\int_{-x_1}^{x_2} f_2(x)\,dx = \int_{-x_1}^{x_2} dx \int_{-\infty}^{+\infty} f_1(\xi)\,k(x,\xi)\,d\xi\,,$$

qui a, d'après (10.2),un sens pour tout x_1 et x_2 fini et positif.

Or, on a d'une part, d'après le théorème de Dini-Lebesgue

$$\int_{-x_1}^{x_2} dx \int_{-\infty}^{\infty} f_1(\xi)k(x,\xi)\,d\xi = \int_{-x_1}^{x_2} dx \lim_{m,n\to\infty} \int_{-m}^{n} f_1(\xi)\,k(x,\xi)\,d\xi =$$

$$= \lim_{m,n\to\infty} \int_{-x_1}^{x_2} dx \int_{-m}^{n} f_1(\xi)\,k(x,\xi)\,d\xi\,,$$

car de

$$f_1(x) \geqslant 0 \quad \text{et} \quad k(x,\xi) \geqslant 0\,,$$

il résulte que

$$0 \leqslant \int_{-m}^{n} f_1(\xi)\,k(x,\xi)\,d\xi \leqslant \int_{-\infty}^{+\infty} f_1(\xi)\,k(x,\xi)\,d\xi = f_2(x).$$

Donc,

$$\int_{-x_1}^{x_2} dx \int_{-\infty}^{+\infty} f_1(\xi)k(x,\xi)\,d\xi = \lim_{m,n\to\infty} \int_{-m}^{n} f_1(\xi)\,d\xi \int_{-x_1}^{x_2} k(x,\xi)\,dx =$$

$$= \int_{-\infty}^{+\infty} f_1(\xi)\,d\xi \int_{-x_1}^{x_2} k(x,\xi)\,dx\,.$$

D'autre part, en tenant compte de

$$\int_{-x_1}^{x_2} k(x,\xi)\,d\xi \leqslant 1\,,$$

$f_1(\xi)$ étant par hypothèse intégrable dans $(-\infty, +\infty)$, par une seconde application du théorème de Dini-Lebesgue, on obtient

$$\lim_{x_1,x_2\to\infty} \int_{-x_1}^{x_2} dx \int_{-\infty}^{+\infty} f_1(\xi)\,k(x,\xi)\,d\xi = \lim_{x_1,x_2\to\infty} \int_{-\infty}^{+\infty} f_1(\xi)\,d\xi \int_{-x_1}^{x_2} k(x,\xi)\,dx =$$

$$= \int_{-\infty}^{\infty} \lim_{x_1,x_2\to\infty} \left\{ f_1(\xi) \int_{-x_1}^{x_2} k(x,\xi)\,dx \right\} d\xi =$$

$$= \int_{-\infty}^{+\infty} f_1(\xi)\,d\xi = 1.$$

ce qui démontre l'affirmation.

Etant donné que le noyau $k(x,\xi)$ est une fonction de deux variables, on peut exprimer d'une infinité de manières $f_2(x)$ au moyen de $f_1(x)$ par la formule (10.1). On pourrait dire que ce noyau représente la loi selon laquelle la population π_1 s'est transformée en π_2. C'est pourquoi nous appellererons cette fonction "la loi de révolution $\pi_1 \to \pi_2$". Il existe évidemment un nombre infini de lois de révolution pour lesquelles la population π_1 pourrait être transformée en population π_2. La spécification de cette loi dépendra de la nature du problème étudié.

Par exemple, il arrive souvent dans la pratique que la population π_1 a subi des changements de façon telle que la valeur du caractère observé ξ soit transformée d'après l'une de deux lois suivantes

$$\text{a)} \quad \xi = x - t, \qquad -\infty < t < +\infty,$$

$$\text{b)} \quad \xi = \tau x, \qquad \tau \geqslant 0,$$

où t et τ représentent dans le cas général des variables aléatoires avec les lois de probabilité $r(t)$ et $\rho(\tau)$. La loi de probabilité de la population transformée aura les formes suivantes

$$\text{a)} \quad f_2(x) = \int_{-\infty}^{\infty} f_1(\xi)\, r(x-\xi)\, d\xi,$$

$$\text{b)} \quad f_2(x) = \int_0^{\pm\infty} f_1(\xi)\, \frac{\xi}{x^2}\, \rho\left(\frac{\xi}{x}\right) d\xi, \qquad x \gtrless 0 .$$

Pour que l'intégrale de la formule b) ait un sens, il faut que l'intégrale

$$\int_1^{\infty} \frac{|\psi(t)|}{t}\, dt, \quad (\psi(t) = \text{f.c. de } \rho(\tau)),$$

converge. Il est facile de constater que les deux noyaux figurant dans les formules a) et b) satisfont aux conditions (10.2) et (10.3).

En général, si un paramètre p d'une loi de probabilité $k(x;p)$ commence à varier en suivant la loi $f_1(p)$, la variable x changera alors sa loi de probabilité d'après (10.1), c'est-à-dire

$$f_2(x) = \int_{-\infty}^{+\infty} f_1(p)\, K(x;p)\, dp .$$

Si nous avons le cas de n paramètres, il sera

$$f_2(x) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_1(p_1,\ldots,p_n)\, K(x;p_1,\ldots,p_n)\, dp_1,\ldots,dp_n.$$

Enfin, nous avons vu que Fisher et Mahalanobis supposent, à propos de la séparation et de la classification des populations, que les deux fonctions f_1 et f_2 sont des lois de Laplace et qu'elles ont les mêmes matrices de dispersion. Ce cas spécial peut être réduit à a) où la loi de probabilité f_1 serait la fonction de Dirac, c'est-à-dire

$$\delta_{(\xi_0)}(\xi) = \begin{cases} \infty, & \xi = \xi_0, \\ 0, & \xi \neq \xi_0 . \end{cases}$$

avec la propriété

$$\int_{-\infty}^{+\infty} \delta_{(\xi_0)}(\xi)\, d\xi = 1 .$$

11. APPROXIMATION EN MOYENNE D'UNE LOI DE RÉVOLUTION

Supposons que dans une école on mesure la taille des enfants agés de 13 ans au commencement de chaque année. A la fin de l'année on mesure de nouveau la taille de ces mêmes enfants. Il est évident que la seconde répartition des tailles ne sera pas identique à la première, vu la croissance sensible des enfants de cet âge. En répétant N fois la même expérience, nous avons reçu N couples de lois de probabilité. Soient $\varphi_i(x)$ et $f_i(x)$, $(i = 1,2,\ldots,N)$, les lois de probabilité de la taille des enfants au commencement, respectivement à la fin de l'année. Il est naturel d'admettre qu'il existe une certaine dépendance entre les lois de probabilité $f_i(x)$ et $\varphi_i(x)$ et supposons qu'elle soit de la forme suivante

$$f(x) = \int_{-\infty}^{+\infty} \varphi(\xi)\, K(x,\xi)\, d\xi . \tag{11.1}$$

Essayons à présent de trouver parmi toutes lois de révolution $k(x,\xi)$ celle qui donnerait la meilleure représentation possible, donc l'estimation, de la loi de probabilité $f(x)$. La détermination de la fonction $k(x,\xi)$ nous permettra d'établir une dépendance statistique entre les lois de probabilité $f(x)$ et $\varphi(x)$. Cette dépendance statistique, à savoir la correlation existant entre les deux ensembles de répartitions de la taille des enfants, peut être plus ou moins forte et nous établirons dans le prochain paragraphe un coefficient représentant une mesure de cette dépendance.

En interprétant l'expression "la meilleure estimation possible" comme l'approximation en moyenne d'une fonction, nous allons déterminer la fonction $K(x,\xi)$ de telle façon que l'expression

$$\sum_{i=1}^{N} \left\{ f_i(x) - \int_{-\infty}^{+\infty} K(x,\xi)\, \varphi_i(\xi)\, d\xi \right\}^2 \tag{11.2}$$

soit minimum. Ce sera le cas si la fonction $k(x,\xi)$ satisfait à l'équation

$$\sum_{i=1}^{N} \varphi_i(\xi) \left\{ f_i(x) - \int_{-\infty}^{+\infty} K(x,t)\, \varphi_i(t)\, dt \right\} = 0 ,$$

ou

$$\sum_{i=1}^{N} \varphi_i(\xi)\, f_i(x) = \int_{-\infty}^{+\infty} K(x,t) \sum_{i=1}^{N} \varphi_i(\xi)\, \varphi_i(t)\, dt , \tag{11.3}$$

c'est-à-dire une équation intégrale singulière de Fredholm de première espèce à noyau dégénéré (de Goursat), symétrique et positif.

Dans la pratique nous rencontrons le plus fréquemment le cas où nous connaissons non pas les lois de probabilité $f_i(x)$ et $\varphi_i(x)$ mais les fréquences par classes d'intervalles, c'est-à-dire les valeurs.

$$f_{i\nu} = \int_{(\nu-1)h}^{\nu h} f_i(x)\, dx , \qquad \varphi_{i\nu} = \int_{(\nu-1)h}^{\nu h} \varphi_i(x)\, dx ,$$

($\nu = 1,2,\ldots,n$, $i = 1,2,\ldots,N$), h étant la longueur de l'intervalle de classe et n le nombre d'intervalles. En posant

(11.4)
$$K_{\nu\zeta} = \int_{(\nu-1)h}^{\nu h} K(x, \xi_\zeta)\, dx$$

avec

$$(\zeta-1)h \leqslant \xi_\zeta \leqslant \zeta h,$$

(11.1) prend la forme de la transformation linéaire

(11.5)
$$f_\nu = \sum_{\zeta=1}^{n} K_{\nu\zeta}\, \varphi_\zeta, \qquad \nu = 1,2,\ldots,n,$$

et la condition (10.4) devient

(11.6)
$$\sum_{\nu=1}^{n} K_{\nu\zeta} = 1.$$

L'équation intégrale (11.3) se réduit alors à un système à n^2 équations linéaires avec les inconnues $K_{\nu\zeta}$ ($\nu,\zeta = 1,2,\ldots,n$)

(11.7)
$$\sum_r K_{\nu r} \sum_i \varphi_{i\zeta}\, \varphi_{ir} = \sum_i f_{i\nu}\, \varphi_{i\zeta}, \quad \nu,\zeta=1,\ldots,n,$$

ou, en désignant

$$S_{\zeta r} = \sum_i \varphi_{i\nu}\, \varphi_{ir}, \quad Q_{\nu\zeta} = \sum_i f_{i\nu}\, \varphi_{i\zeta}$$

le système (11.7) devient

(11.8)
$$S_{\zeta 1} k_{\nu 1} + S_{\zeta 2} k_{\nu 2} + \ldots + S_{\zeta n} k_{\nu n} = Q_{\nu\zeta},$$

($\nu,\zeta = 1,2,\ldots,n$), qui se décompose en n systèmes de n équations linéaires

(11.9)
$$\begin{bmatrix} S_{11} & S_{12} & \ldots & S_{1n} \\ S_{21} & S_{22} & & S_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ S_{n1} & S_{n2} & & S_{nn} \end{bmatrix} \begin{bmatrix} k_{\nu 1} \\ k_{\nu 2} \\ \ldots \\ k_{\nu n} \end{bmatrix} = \begin{bmatrix} Q_{\nu 1} \\ Q_{\nu 2} \\ \ldots \\ Q_{\nu n} \end{bmatrix}$$

$\nu = 1,2,\ldots,n$

Pour que ces systèmes soient résolubles, il faut que le déterminant S associé aux systèmes (11.9), c'est-à-dire

$$S = \begin{vmatrix} \sum_{i=1}^{N} \varphi_{i1}^2 & \sum_{i=1}^{N} \varphi_{i1}\varphi_{i2} & \ldots & \sum_{i=1}^{N} \varphi_{i1}\varphi_{in} \\ \sum_{i=1}^{N} \varphi_{i1}\varphi_{i2} & \sum_{i=1}^{N} \varphi_{i2}^2 & \ldots & \sum_{i=1}^{N} \varphi_{i2}\varphi_{in} \\ \ldots & \ldots & \ldots & \ldots \\ \sum_{i=1}^{N} \varphi_{i1}\varphi_{in} & \sum_{i=1}^{N} \varphi_{i2}\varphi_{in} & \ldots & \sum_{i=1}^{N} \varphi_{in}^2 \end{vmatrix}$$

soit différent de zéro. En décomposant ce déterminant en une somme de déterminants n'ayant que des monomes comme éléments nous constatons que chaque déterminant, ainsi formé, aura au moins deux colonnes identiques si $N < n$. Par conséquent, il faut que $N \geqslant n$ pour que le déterminant S, associé aux systèmes (11.9) soit différent de zéro.

Les solutions des systèmes (11.9) sont données par les formules

$$k_{\nu\zeta} = S^{1\zeta} Q_{\nu 1} + S^{2\zeta} Q_2 + \ldots + S^{n\zeta} Q_{\nu n}, \qquad \nu, \zeta = 1, 2, \ldots, n, \tag{11.10}$$

où les coefficients $S^{\nu\zeta}$ désignent des éléments de la matrice inverse de $\{S_{\nu\zeta}\}$.

Il est facile de vérifier que les coefficients (11.10) satisfont les conditions (11.6).

En effet, les égalités

$$\sum_\nu Q_{\nu\zeta} = \sum_\nu \sum_i f_{i\nu} \varphi_{i\zeta} = \sum_i \varphi_{i\zeta} \sum_\nu f_{i\nu} = \sum_i \varphi_{i\zeta},$$

$$\sum_j S_{\zeta j} = \sum_j \sum_i \varphi_{i\zeta} \varphi_{ij} = \sum_i \varphi_{i\zeta} \sum_j \varphi_{ij} = \sum_i \varphi_{i\zeta},$$

nous montrent que

$$\sum_\nu Q_{\nu\zeta} = \sum_j S_{\zeta j},$$

d'où

$$\begin{aligned}\sum_\nu k_{\nu\zeta} &= \sum_\nu (S^{1\zeta} Q_{\nu 1} + \ldots + S^{n\zeta} Q_{\nu n}) = \\ &= S^{1\zeta} \sum_j S_{1j} + \ldots + S^{n\zeta} \sum_j S_{nj} = \\ &= \sum_j (S^{1\zeta} S_{1j} + \ldots + S^{n\zeta} S_{nj}) = \\ &= S^{1\zeta} S_{1\zeta} + \ldots + S^{n\zeta} S_{n\zeta} = \frac{1}{S} S,\end{aligned}$$

c'est-à-dire

$$\sum_\nu k_{\nu\zeta} = 1.$$

Remarquons que la conditions (10.3) n'est pas nécessaire et que, par conséquent, les fréquences f_ν peuvent conserver leurs propriétés dans le cas où quelques-uns des nombres k_ν sont négatifs.

Enfin, la somme totale des différences entre les valeurs exactes $f_{i\nu}$ et les valeurs ajustées, obtenues par la formule (11.5), est égale à zéro. En effet,

$$\begin{aligned}&\sum_{i=1}^N \sum_{\nu=1}^n (f_{i\nu} - \sum_{\zeta=1}^n k_{\nu\zeta} \varphi_{i\zeta}) = \\ &= \sum_{i=1}^N \sum_{\nu=1}^n f_{i\nu} - \sum_{i=1}^N \sum_{\zeta=1}^n \varphi_{i\zeta} \sum_{\nu=1}^n k_{\nu\zeta} = \\ &= N - N = 0.\end{aligned}$$

Etant donné que

$$\sum_i \sum_\nu \sum_\zeta k_{\nu\zeta} \varphi_{i\zeta} (f_{i\nu} - \sum_j k_{\nu j} \varphi_{ij}) =$$

$$= \sum_\nu \sum_\zeta k_{\nu\zeta} (Q_{\nu\zeta} - \sum_j k_{\nu j} S_{\zeta j}) =$$

$$= \sum_\nu \sum_\zeta k_{\nu\zeta} (Q_{\nu\zeta} - Q_{\nu\zeta}) = 0 ,$$

la variance résiduelle sera donnée par

$$= \frac{1}{nN} \sum_{i=1}^{N} \sum_{\nu=1}^{n} (f_{i\nu} - \sum_{\zeta=1}^{n} k_{\nu\zeta} \varphi_{i\zeta})^2 =$$

$$= \frac{1}{nN} \sum_{i=1}^{N} \sum_{\nu=1}^{n} f_{i\nu} (f_{i\nu} - \sum_{\zeta=1}^{n} k_{\nu\zeta} \varphi_{i\zeta}) .$$

Si nous obtenons comme solutions des systèmes (11.9) $k_{\nu\nu} = 1$, ($\nu = 1, 2, \ldots, n$) et $k_{\nu\zeta} = 0$ ($\nu \neq \zeta$), les deux répartitions de fréquences seront identiques ; $f_\nu = \varphi_\nu$ ($\nu = 1, 2, \ldots, n$). Dans notre exemple un pareil cas n'aura certainement pas lieu, étant donné que les acroissements de la taille des enfants à 13 ans ne seront jamais tous égaux à zéro.

Considérons à présent le cas inverse, le cas où les fréquences φ_ν sont représentées à l'aide d'une transformation linéaire, par les fréquences f_ν , c'est-à-dire

$$\text{(11.11)} \qquad \varphi_\nu = \sum_{\zeta=1}^{n} k^{\zeta\nu} f_\zeta \ , \quad \nu = 1, 2, \ldots, n ,$$

où ces coefficients $k^{\zeta\nu}$ définissent la loi de révolution $\pi_f \to \pi_\varphi$ et dont les valeurs seront représentées par les solutions de n systèmes à n équations linéaires suivantes

$$\text{(11.12)} \qquad \begin{bmatrix} P_{11} & P_{12} & \ldots & P_{1n} \\ P_{21} & P_{22} & \ldots & P_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ P_{n1} & P_{n2} & \ldots & P_{nn} \end{bmatrix} \begin{bmatrix} k^{1\nu} \\ k^{2\nu} \\ \ldots \\ k^{n\nu} \end{bmatrix} = \begin{bmatrix} Q_{1\nu} \\ Q_{2\nu} \\ \ldots \\ Q_{n\nu} \end{bmatrix}$$

où nous avons posé

$$P_{\nu\zeta} = \sum_{i=1}^{N} f_{i\nu} f_{i\zeta} \ , \qquad \nu, \zeta = 1, 2, \ldots, n .$$

Les solutions des systèmes (11.12) sont données par les formules

$$\text{(11.13)} \quad k^{\zeta\nu} = P^{1\zeta} Q_{1\nu} + P^{2\zeta} Q_{2\nu} + \ldots + P^{n\zeta} Q_{n\nu} \ , \qquad \nu, \zeta = 1, 2, \ldots, n ,$$

où les coefficients $P^{\nu\zeta}$ désignent les éléments de la matrice inverse de $(P_{\nu\zeta})$. Les coefficients (11.13) satisfont les conditions

$$\text{(11.14)} \qquad \sum_{\nu=1}^{n} k^{\zeta\nu} = 1 \ , \quad \zeta = 1, 2, \ldots, n .$$

Remarquons que les deux déterminants symétriques S et P ne peuvent pas être négatifs. Afin de démontrer cette propriété considérons par exemple le déterminant S. L'espérance mathématique de l'expression

$$(\sum_{i=1}^{n} t_i \varphi_i)^2$$

c'est-à-dire

$$E(\sum_{i=1}^{n} t_i \varphi_i)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_{ij} t_i t_j ,$$

où les coefficients λ_{ij} représentent les moments du deuxième ordre des variables φ_i, ne peut pas être négative et le deuxième membre représente, par conséquent, une forme quadratique non-négative par rapport aux $t_1, \dots, t_n$ avec la matrice des moments

$$\begin{bmatrix} \lambda_{11} & \dots & \lambda_{1n} \\ \dots & \dots & \dots \\ \lambda_{n1} & \dots & \lambda_{nn} \end{bmatrix} = N^{-n} \begin{bmatrix} S_{11} & \dots & S_{1n} \\ \dots & \dots & \dots \\ S_{n1} & \dots & S_{nn} \end{bmatrix}$$

Il en résulte que la matrice symétrique $(S_{\nu\tau})$ est non-négative et que tous les nombres caractéristiques sont eux aussi non-négatifs. Etant donné que le déterminant S est égal au produit de ces nombres caractéristiques, on en conclut que $S \geqslant 0$, c.q.f.d.

Exemple (11.1). - En comparant les Tables de mortalité de différentes époques on constate que la répartition des décédés par rapport à leurs âges évolue très sensiblement au cours du temps. Le progrès des sciences et l'élévation du niveau de vie ont amené une grande amélioration dans la mortalité au cours de derniers 150 ans. Prenons comme exemple les répartitions des décédés de la ville de Bâle pour la période 1901-1950. Etant donné que cette ville était pendant cette période épargnée des guerres et d'autres cataclysmes, il s'agit ici d'une évolution normale. Les Tableaux 11.1 et 11.2 donnent les répartitions des fréquences moyennes, absolues et relatives respectivement pour les périodes 1901-1905, 1906-1910, ..., 1946-1950 (39).

TABLEAU 11.1 DECEDES PAR GROUPE D'AGE (Nombres absolus)

AGE	0-9	10-19	20-29	30-39	40-49	50-59	60-69	de 70	Total
1901/05	608	57	114	126	132	185	232	272	1726
1906/10	433	62	106	131	159	193	254	311	1649
1911/15	319	68	89	126	161	209	262	373	1907
1916/20	178	92	153	143	181	249	290	426	1712
1921/25	140	53	92	87	156	253	298	447	1526
1926/30	120	40	104	87	139	261	350	506	1607
1931/35	122	31	105	109	148	282	405	553	1755
1936/40	103	27	67	95	145	259	463	723	1882
1941/45	120	18	55	76	151	227	443	764	1854
1946/50	116	18	40	63	141	261	442	957	2038

TABLEAU 11.2 DECEDES PAR GROUPE D'AGE (EN %)

âge	0-9	10-19	20-29	30-39	40-49	50-59	60-69	au dessus de 70
1901/05	352	33	66	73	76	107	135	158
1906/10	263	38	64	79	96	117	154	189
1911/15	199	42	55	78	100	130	163	233
1916/20	104	54	89	84	106	145	169	249
1921/25	92	35	60	57	102	166	195	293
1926/30	75	25	65	54	86	162	218	315
1931/35	70	18	60	62	84	161	230	315
1936/40	55	14	36	50	77	138	246	384
1941/45	65	10	30	41	81	122	239	412
1946/50	57	9	20	31	69	128	216	470

Afin de déterminer les valeurs $k_{\nu\zeta}$ $(\nu,\zeta = 1,2,\ldots,8)$ d'après la méthode décrite ci-dessus, posons

$$f_{i+1,\nu} = \varphi_{i\nu} \qquad \begin{matrix} i = 1,2,\ldots,10\,, \\ \nu = 1,2,\ldots,8\,. \end{matrix}$$

Les matrices $(Q_{\nu\zeta})$ et $(S_{\nu\zeta})$ seront alors

$$(Q_{\nu\zeta}) = \begin{bmatrix} 198457 & 44573 & 76894 & 88354 & 120336 & 173162 & 232699 & 340525 \\ 32422 & 8815 & 16223 & 17401 & 25248 & 39022 & 52218 & 77651 \\ 60402 & 15421 & 29619 & 32575 & 47870 & 75488 & 105120 & 158505 \\ 67812 & 17378 & 32496 & 36116 & 52835 & 81897 & 114585 & 174881 \\ 87156 & 22803 & 44148 & 48642 & 72598 & 115192 & 163501 & 253960 \\ 126853 & 32873 & 66439 & 73262 & 110419 & 178293 & 258182 & 401679 \\ 170799 & 42945 & 87952 & 99043 & 152085 & 246536 & 366042 & 583598 \\ 236099 & 60129 & 125229 & 140607 & 219609 & 359410 & 537653 & 869201 \end{bmatrix}$$

$$(S_{\nu\zeta}) = \begin{bmatrix} 149074 & 32256 & 56698 & 65851 & 91319 & 133229 & 182282 & 269291 \\ & 8675 & 15337 & 16664 & 23313 & 34798 & 46017 & 67940 \\ & & 29063 & 31122 & 44410 & 68881 & 93878 & 139611 \\ & & & 34532 & 49264 & 75625 & 105018 & 157924 \\ & & & & 72559 & 113330 & 160161 & 246644 \\ & & & & & 181647 & 259183 & 402307 \\ & & & & & & 381588 & 601873 \\ & & & & & & & 974410 \end{bmatrix}$$

Les solutions des systèmes (11.9) sònt (1)

K_{11}	= 1,295124	K_{21}	= 0,009704	K_{31}	= 0,048601
K_{12}	= 3,559441	K_{22}	= 0,847320	K_{32}	= 0,889007
K_{13}	= 0,693264	K_{23}	= - 0,235984	K_{33}	= - 0,913618
K_{14}	= 0,415440	K_{24}	= - 0,732079	K_{34}	= - 0,392112
K_{15}	= - 2,449983	K_{25}	= 0,472662	K_{35}	= 0,203077
K_{16}	= - 0,462046	K_{26}	= 0,321178	K_{36}	= 0,727505
K_{17}	= 0,302118	K_{27}	= 0,082910	K_{37}	= 0,344997
K_{18}	= 0,201001	K_{28}	= - 0,133068	K_{38}	= - 0,283165
K_{41}	= 0,017792	K_{51}	= 0,053443	K_{61}	= 0,050822
K_{42}	= 1,696071	K_{52}	= 1,814171	K_{62}	= 1,673553
K_{43}	= - 0,778826	K_{53}	= - 0,336988	K_{63}	= - 0,024786
K_{44}	= - 0,336155	K_{54}	= - 0,567827	K_{64}	= - 0,552959
K_{45}	= 0,323537	K_{55}	= - 0,345699	K_{65}	= - 1,052749
K_{46}	= 0,147964	K_{56}	= 0,497367	K_{66}	= 0,468044
K_{47}	= 0,383125	K_{57}	= 0,268409	K_{67}	= 1,020751
K_{48}	= - 0,157265	K_{58}	= - 0 ,023957	K_{68}	= - 0,182600
K_{71}	= - 0,151234	K_{81}	= -0,324253		
K_{72}	= - 2,398636	K_{82}	= -7,080928		
K_{73}	= 0,333762	K_{83}	= 2,263176		
K_{74}	= 1,443308	K_{84}	= 1,722383		
K_{75}	= 0,884617	K_{85}	= 2,964539		
K_{76}	= - 0,277144	K_{86}	= -0,422868		
K_{77}	= 0,041784	K_{87}	= -1,444095		
K_{78}	= 0,390924	K_{88}	= 1,188131		

$$f_{i\nu} = \sum_{\zeta=1}^{8} k_{\nu\zeta}\, f_{i+1,\zeta}$$

$$\nu = 1, 2, \dots, 8 .$$

(1) Le calcul a été effectué par M. H. Rutishauser (E. T. H. Institut für angewandte Mathematik, Zürich).

Le tableau 11.3 donne les valeurs ajustées de f_ν, obtenues par la formule (11.5).

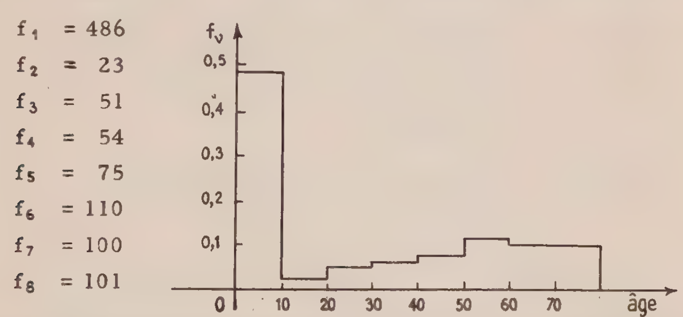
TABLEAU 11.3

âge	0-9	10-19	20-29	30-39	40-49	50-59	60-69	au dessus de 70
1901/05	349	32	62	70	78	108	137	164
1906/10	269	39	71	83	92	115	151	180
1911/15	198	42	53	77	101	130	164	235
1916/20	100	53	84	82	109	146	171	255
1921/25	97	36	67	61	98	164	192	285
1926/30	73	25	65	54	86	162	218	315
1931/35	66	17	55	59	87	162	233	321
1936/40	58	14	40	52	75	137	244	380
1941/45	64	10	29	40	82	122	240	413

Il est facile de vérifier que la somme des différences des éléments correspondants des Tableaux, 11.2 et 11.3 est égale à zéro et que l'écart-type de ces différences ne dépasse pas 3,3 ‰.

On pourrait maintenant reconstituer la répartition moyenne des décédés à Bâle pour la période 1896-1900. En utilisant la répartition pour la période 1901-1905 et les nombres trouvés $k_{\nu\zeta}$, on obtient par la formule (11.5) les valeurs

$f_1 = 486$
$f_2 = 23$
$f_3 = 51$
$f_4 = 54$
$f_5 = 75$
$f_6 = 110$
$f_7 = 100$
$f_8 = 101$



qui représentent l'estimation de la répartition moyenne des fréquences relatives des décédés à Bâle pour la période 1896-1900

12. COEFFICIENT DE CORRÉLATION DE DEUX POPULATIONS

Dans le but de former une mesure de l'interdépendance de deux ensembles de populations π_{i1} et π_{i2} $(i = 1,2,\ldots,N)$, posons

$$R = \frac{Q}{\sqrt{PS}}, \qquad (12.1)$$

où l'on désigne de nouveau

$$Q = \left| \sum_{i=1}^{N} \varphi_{i\nu} f_{i\zeta} \right| , \quad P = \left| \sum_{i=1}^{N} f_{i\nu} f_{i\zeta} \right| , \quad S = \left| \sum_{i=1}^{N} \varphi_{i\nu} \varphi_{i\zeta} \right| .$$

Le coefficient R ainsi défini sera appelé le coefficient de corrélation de π_1 et de π_2. Dans le cas n = 2, R devient le coefficient de la corrélation ordinaire de deux variables aléatoires, c'est-à-dire

$$R = r = \frac{\begin{vmatrix} \sum_{i=1}^{N} \varphi_{i1} f_{i1} & \sum_{i=1}^{N} \varphi_{i1} \\ \sum_{i=1}^{N} f_{i1} & N \end{vmatrix}}{\begin{vmatrix} \sum_{i=1}^{N} f_{i1}^2 & \sum_{i=1}^{N} f_{i1} \\ \sum_{i=1}^{N} f_{i1} & N \end{vmatrix}^{1/2} \cdot \begin{vmatrix} \sum_{i=1}^{N} \varphi_{i1}^2 & \sum_{i=1}^{N} \varphi_{i1} \\ \sum_{i=1}^{N} \varphi_{i1} & N \end{vmatrix}^{1/2}} =$$

$$= \frac{N \sum_{i=1}^{N} \varphi_{i1} f_{i1} - \sum_{i=1}^{N} \varphi_{i1} \sum_{i=1}^{N} f_{i1}}{\left\{ N \sum_{i=1}^{N} \varphi_{i1}^2 - \left(\sum_{i=1}^{N} \varphi_{i1}\right)^2 \right\}^{1/2} \left\{ N \sum_{i=1}^{N} f_{i1}^2 - \left(\sum_{i=1}^{N} f_{i1}\right)^2 \right\}^{1/2}} .$$

Démontrons que $|R|$ est compris entre 0 et 1. En effet, nous pouvons transformer le déterminant Q de la façon suivante

$$Q = \begin{vmatrix} \sum_{i=1}^{N} \varphi_{i1} f_{i1} & \sum_{i=1}^{N} \varphi_{i1} f_{i2} & \dots & \sum_{i=1}^{N} \varphi_{i1} f_{in} \\ \sum_{i=1}^{N} \varphi_{i2} f_{i1} & \sum_{i=1}^{N} \varphi_{i2} f_{i2} & \dots & \sum_{i=1}^{N} \varphi_{i2} f_{in} \\ \dots & \dots & \dots & \dots \\ \sum_{i=1}^{N} \varphi_{in} f_{i1} & \sum_{i=1}^{N} \varphi_{in} f_{i2} & \dots & \sum_{i=1}^{N} \varphi_{in} f_{in} \end{vmatrix}$$

$$= \sum_{1 \leq i_1 < i_2 \dots < i_n \leq N} \begin{vmatrix} f_{i_1 1} & f_{i_2 1} & \dots & f_{i_n 1} \\ f_{i_1 2} & f_{i_2 2} & \dots & f_{i_n 2} \\ \dots & \dots & \dots & \dots \\ f_{i_1 n} & f_{i_2 n} & \dots & f_{i_n n} \end{vmatrix} \begin{vmatrix} \varphi_{i_1 1} & \varphi_{i_2 1} & \dots & \varphi_{i_n 1} \\ \varphi_{i_1 2} & \varphi_{i_2 2} & \dots & \varphi_{i_n 2} \\ \dots & \dots & \dots & \dots \\ \varphi_{i_1 n} & \varphi_{i_2 n} & \dots & \varphi_{i_n n} \end{vmatrix}$$

En posant, pour simplifier,

$$(12.2) \quad \Delta^{f}_{i_1 i_2 \dots i_n} = \begin{vmatrix} f_{i_1 1} \dots f_{i_n 1} \\ \dots\dots \\ f_{i_1 n} \dots f_{i_n n} \end{vmatrix}$$

et

(12.3)
$$\Delta^{\varphi}_{i_1 i_2 \dots i_n} = \begin{vmatrix} \varphi_{i_1 1} & \cdots & \varphi_{i_n 1} \\ \dots & \dots & \dots \\ \varphi_{i_1 n} & \cdots & \varphi_{i_n n} \end{vmatrix}$$

on obtient finalement

(12.4)
$$Q = \frac{1}{n!} \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \dots \sum_{i_n=1}^{N} \Delta^{\varphi}_{i_1 i_2 \dots i_n} \Delta^{f}_{i_1 i_2 \dots i_n},$$

où les indices $i_1, \dots i_n$ prennent indépendamment toutes les valeurs entières $1, 2, \dots, N$. A chaque système de n entiers $(i_1, i_2, \dots, i_n)$ correspond un terme de la somme. On peut écrire plus simplement cette formule en numérotant les systèmes d'entiers $(i_1, i_2, \dots, i_n)$. On donnera le numéro $i = 1$ au système dont

$$\sum_{j=1}^{n} i_j = n$$

puis les numéros $i = 2, 3, \dots, n+1$ aux systèmes dont

$$\sum_{j=1}^{n} i_j = n+1$$

et ainsi de suite, Ainsi, à chaque système distinct de $(i_1, i_2, \dots, i_n)$ on aura associé un seul entier $i = i(i_1, i_2, \dots, i_n)$, $i = 1; 2, \dots$. Le déterminant Q s'écrit alors sous la forme condensée

(12.5)
$$Q = \frac{1}{n!} \sum_i \Delta^{\varphi}_i \Delta^{f}_i .$$

D'une façon analogue on peut écrire les déterminants P et S

(12.6)
$$P = \frac{1}{n!} \sum_i (\Delta^{f}_i)^2 ,$$

(12.7)
$$S = \frac{1}{n!} \sum_i (\Delta^{\varphi}_i)^2 .$$

L'application immédiate de l'inégalité de Schwartz

$$\left(\frac{1}{n!} \sum_i \Delta^{\varphi}_i \Delta^{f}_i\right)^2 - \frac{1}{n!} \sum_i (\Delta^{\varphi}_i)^2 \frac{1}{n!} \sum_i (\Delta^{f}_i)^2 \leqslant 0$$

nous montre alors que $R^2 \leqslant 1$, c.q.f.d.

L'égalité $R^2 = 1$ aura lieu si

$$\frac{Q}{P} = \frac{S}{Q} = \frac{1}{\omega}$$

On en tire

$$\omega Q - P = 0$$

$$\omega S - Q = 0$$

ou, en utilisant les formules (12.5), (12.6) et 12.7), on aura

$$\omega Q - P = \frac{1}{n!} \sum_i \Delta_i^f \; (\omega \Delta_i^\varphi - \Delta_i^f) = 0 \, ,$$

$$\omega S - Q = \frac{1}{n!} \sum_i \Delta_i^\varphi \; (\omega \Delta_i^\varphi - \Delta_i^f) = 0 \, .$$

En multipliant la deuxième égalité par ω et en la retranchant de la première il vient

$$\sum_i (\Delta_i^f - \omega \Delta_i^\varphi)^2 = 0 \, .$$

L'expression du premier nombre étant nécessairement non-négative, elle ne peut être nulle que si

$$\Delta_i^f = \omega \Delta_i^\varphi$$

pour chaque système d'indices $i = i(i_1, i_2, \ldots, i_n)$. Or, ce ne sera le cas que si les fréquence f_ν sont des transformations linéaires des φ_ν, c'est-à-dire

$$f_{i_s \nu} = \sum_{\zeta=1}^{n} k_{\nu\zeta} \; \varphi_{i_s \zeta} \qquad \begin{array}{l} \nu = 1, \ldots, n, \\ i_s = 1, \ldots, N, \end{array}$$

ce qui donne

$$\Delta_i^f = \begin{vmatrix} k_{11} & \ldots & k_{1n} \\ \ldots & \ldots & \ldots \\ k_{n1} & \ldots & k_{nn} \end{vmatrix} \Delta_i^\varphi$$

Si f_ν et φ_ζ $(\nu, \zeta = 1, 2, \ldots, n)$ sont indépendantes, $R = 0$, mais ici aussi R peut être nul sans que f_ν et φ_ζ soient indépendantes.

Remarquons que dans le cas où la répartition est continue, le déterminant Q prendra la forme

$$Q = \begin{vmatrix} \int_{-\infty}^{+\infty} \varphi_1(x) \, f_1(x) \, dx & \int_{-\infty}^{+\infty} \varphi_1(x) \, f_2(x) \, dx \ldots & \int_{-\infty}^{+\infty} \varphi_1(x) \, f_n(x) \, dx \\ \int_{-\infty}^{+\infty} \varphi_2(x) \, f_1(x) \, dx & \int_{-\infty}^{+\infty} \varphi_2(x) \, f_2(x) \, dx \ldots & \int_{-\infty}^{+\infty} \varphi_2(x) \, f_n(x) \, dx \\ \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \int_{-\infty}^{+\infty} \varphi_n(x) \, f_1(x) \, dx & \int_{-\infty}^{+\infty} \varphi_n(x) \, f_2(x) \, dx \ldots & \int_{-\infty}^{+\infty} \varphi_n(x) \, f_n(x) \, dx \end{vmatrix} =$$

$$= \frac{1}{n!}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\begin{vmatrix} f_1(x_1)\, f_1(x_2)\ldots f_1(x_n)\\ f_2(x_1)\, f_2(x_2)\ldots f_2(x_n)\\ \ldots\ldots\ldots\ldots\\ f_n(x_1)\, f_n(x_2)\ldots f_n(x_n)\end{vmatrix}\begin{vmatrix} \varphi_1(x_1)\,\varphi_1(x_2)\ldots\varphi_1(x_n)\\ \varphi_2(x_1)\,\varphi_2(x_2)\ldots\varphi_2(x_n)\\ \ldots\ldots\ldots\ldots\\ \varphi_n(x_1)\,\varphi_n(x_2)\ldots\varphi_n(x_n)\end{vmatrix} dx_1\ldots dx_n$$

En remplaçant respectivement $\varphi_i(x)$ et $f_i(x)$ par $f_i(x)$ et $\varphi_i(x)$, on obtient les expressions correspondantes de P et S.

13. DIVISION D'UNE POPULATION EN DEUX SOUS-POPULATIONS

Nous avons vu de quelle façon il a fallu diviser l'espace en deux sous-espaces R_1^k et R_2^k afin de pouvoir entreprendre la classification des éléments particuliers de la population totale $\pi = \pi_1 + \pi_2$, par rapport aux populations π_1 et π_2. Nous avons toujours supposé que les masses des populations π_1 et π_2 étaient déjà séparées et que leurs lois de probabilité $f_1(x)$ et $f_2(x)$ nous étaient connues. Cependant, il se présente souvent le cas où nous connaissons la loi de probabilité de la population π, sans connaître celles de π_1 et π_2. Par conséquent, le problème général de la classification présente deux étapes. La première étape consiste à partager la population totale en deux sous-populations d'après certaines indications où suppositions. Une fois ce partage effectué, on procède à la seconde étape, c'est-à-dire à la division de l'espace R^k en deux sous-espaces.

Le point de départ de K Pearson (37) dans la division des masses était la supposition que les caractères des éléments de deux populations suivent des lois de Laplace, réduisant ainsi tout le problème à la détermination des moments et des coefficients de pondération des populations π_1 et π_2 en vertu des moments connus de la population totale π.

Dans le cas à une dimension les paramètres cherchés $m_1, m_2, \sigma_1, \sigma_2$ et $p = 1-q$ dépendraient alors de la racine bien choisie d'une équation nonique, construite à partir des cinq premiers moments de la population π.

Nous nous proposons de considérer le cas où la forme des lois de probabilité $f_1(x)$ et $f_2(x)$ est inconnue, supposant seulement la connaissance de la loi de probabilité $\varphi(x)$ de la population totale π et de la loi de révolution $\pi_1 \to \pi_2$, c'est-à-dire la fonction $K(x,\xi)$, telle que

$$(13.1)\qquad f_2(x) = \int_{-\infty}^{+\infty} f_1(\xi)\, k(x,\xi)\, d\xi\,.$$

Les fonctions $f_1(x)$ et $f_2(x)$ doivent satisfaire la relation

$$(13.2)\qquad pf_1(x) + qf_2(x) = \varphi(x)$$

et en utilisant (13.1) on obtient

$$(13.3)\qquad f_1(x) - \frac{p-1}{p}\int_{-\infty}^{+\infty} f_1(\xi)\, k(x,\xi)\, d\xi = \frac{1}{p}\varphi(x)\,,$$

c'est-à-dire une équation intégrale singulière de Fredholm de seconde espèce avec le paramètre $\lambda = (p-1)\, p \leqslant 0$ et la fonction inconnue $f_1(x)$.

Avant de résoudre l'équation (13.3) prouvons le théorème suivant

Théorème I. Le noyau itéré d'un noyau représentant une loi de révolution est une loi de révolution.

Afin de démontrer ce théorème utilisons la formule de récurrence pour le $\nu+1$-ième itéré $k_{\nu+1}(x,\xi)$ du noyau $k_1 = k(x,\xi)$

$$k_{\nu+1}(x,\xi) = \int_{-\infty}^{+\infty} k(x,t)\, k_\nu(t,\xi)\, dt. \tag{13.4}$$

Le théorème est vrai pour $\nu = 1$. Dans ce cas nous avons

$$k_2(x,\xi) = \int_{-\infty}^{+\infty} k(x,t)\, k(t,\xi)\, dt.$$

Vu la propriètè (10.3) du noyau $k(x,\xi)$, il est clair que le noyau itéré $k_2(x,\xi)$ sera lui aussi positif. De même

$$\int_{-\infty}^{+\infty} k_2(x,\xi)\, dx = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} k(x,t)\, k(t,\xi)\, dt\, dx .$$

D'après les hypothèses générales sur le noyau $k(x,\xi)$, on peut échanger l'ordre des intégrations figurant au dernier membre. Il vient ainsi

$$\int_{-\infty}^{+\infty} k_2(x,\xi)\, dx = \int_{-\infty}^{+\infty} k(t,\xi)\, dt \int_{-\infty}^{+\infty} k(x,t)\, dx .$$

Vu la propriété (10.4) du noyau $k(x,\xi)$ on obtient finalement

$$\int_{-\infty}^{+\infty} k_2(x,\xi)\, dx = 1 .$$

Par conséquent le noyau itéré $k_2(x,\xi)$ représente lui aussi une loi de révolution. Montrons maintenant que, si le théorème est vrai pour ν, il sera vrai pour $\nu+1$. Dans ce cas on suppose que la fonction $k_\nu(x,\xi)$ satisfait les conditions

$$k_\nu(x,\xi) \geqslant 0 \quad , \qquad \int_{-\infty}^{+\infty} k_\nu(x,\xi)\, dx = 1$$

et on démontre d'une façon analogue, comme pour $\nu = 1$, que le noyau itéré $k_{\nu+1}(x,\xi)$, donné par la formule de récurrence (13.4) représente lui aussi une loi de révolution.

Corollaire I. Si la fonction f x) représente une loi de probabilité, la fonction

$$\int_{-\infty}^{+\infty} k_\nu(x,t)\, f(t)\, dt$$

représente elle aussi une loi de probabilité.

Corollaire II. En désignant la loi de probabilité itérée par

$$f_\nu(x) = \int_{-\infty}^{+\infty} k_1(x,t)\, f_{\nu-1}(t)\, dt ,$$

nous pouvons écrire l'identité

$$\int_{-\infty}^{+\infty} k_\nu(x,t) f_1(t)\, dt = \int_{-\infty}^{+\infty} k_1(x,t) f_\nu(t)\, dt .$$

La démonstration est immédiate.

Considérons à présent notre équation intégrale (13.3). En substituant la fonction $f_1(\xi)$ figurant dans l'intégrale du premier membre par

$$f_1(\xi) = \frac{1}{p}\varphi(\xi) + \lambda \int_{-\infty}^{+\infty} f_1(\xi_1)\, k(\xi, \xi_1)\, d\xi_1$$

et répétant indéfiniment ce même procédé, la fonction cherchée $f_1(x)$ sera donnée par la série

$$f_1(x) = \frac{1}{p}\sum_{\nu=0}^{\infty} \lambda^\nu \int_{-\infty}^{+\infty} \varphi(\xi)\, k_\nu(x,\xi)\, d\xi$$

ou, vu le corollaire II du Théorème I,

$$f_1(x) = \frac{1}{p}\sum_{\nu=0}^{\infty} \lambda^\nu \varphi_{\nu+1}(x) . \tag{13.5}$$

S'il existe un nombre positif M, tel que, quelque soit ν et quelque soit x sur $(-\infty, +\infty)$, on ait

$$\varphi_\nu(x) < M ,$$

la série qui figure dans le second membre de (13.5) sera convergente si $|\lambda| < 1$, a savoir $p > \frac{1}{2}$.

D'autre part le paramètre p doit encore satisfaire les conditions pour que la fonction $f_1(x)$ représente elle aussi une loi de probabilité. On vérifie immédiatement que la masse totale correspondant à la fonction $f_1(x)$ est égale à 1,

$$\int_{-\infty}^{+\infty} f_1(x)\, dx = \frac{1}{p}\sum_{\nu=0}^{\infty} \lambda^\nu \int_{-\infty}^{+\infty} \varphi_{\nu+1}(x)\, dx =$$

$$= \frac{1}{p}\sum_{\nu=0}^{\infty} \lambda^\nu = \frac{1}{p}\ \frac{1}{1-\lambda} = 1 .$$

La solution $f_1(x)$ doit être positive ; on constate alors que, pour une loi de probabilité donnée $\varphi(x)$ de la population totale π et pour une loi de révolution donnée $k(x,\xi)$, les proportions des masses des populations π_1 et π_2 ne peuvent pas être arbitraires. Leurs intervalles de variations dépendent essentiellement de $k(x,\xi)$ et de $\varphi(x)$.

Si nous connaissons les fréquences par classe d'intervalles au lieu de la loi de probabilité, de la population c'est-à-dire les valeurs

$$\varphi_\nu = \int_{(\nu-1)h}^{\nu h} \varphi(x)\, dx , \qquad \nu = 1, \dots, n ,$$

h étant la longueur de l'intervalle de classe et n le nombre d'intervalles et si nous désignons par

$$f_{1\nu} = \int_{(\nu-1)h}^{\nu h} f_1(x)\,dx\,, \qquad \nu = 1,\ldots,n\,,$$

$$k_{\nu\zeta} = \int_{(\nu-1)h}^{\nu h} k(x,\xi_\zeta)\,dx\,, \quad \nu,\zeta = 1,\ldots,n\,,$$

avec

$$(\zeta-1)\,h \leqslant \xi_\zeta \leqslant \zeta h$$

l'équation intégrale (13.3) se réduit alors à un système de n équations linéaires avec les inconnues $f_{1\nu}$ ($\nu = 1,2,\ldots,n$)

$$pf_{1\nu} + q\sum_{\zeta=1}^{n} f_{1\zeta}\,k_{\nu\zeta} = \varphi_\nu\,, \qquad \nu = 1,\ldots,n\,, \tag{13.6}$$

ou bien

$$\begin{bmatrix} p/q + k_{11} & k_{12} & \ldots & k_{1n} \\ k_{21} & p/q + k_{22} & \ldots & k_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ k_{n1} & k_{n2} & \ldots & p/q + k_{nn} \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ \ldots \\ f_{1n} \end{bmatrix} = \frac{1}{q} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \ldots \\ \varphi_n \end{bmatrix} \tag{13.7}$$

Vu que les valeurs cherchées $f_{1\nu}$ ($\nu = 1,2,\ldots,n$) représentent les fréquences des masses de la population π_1, elles doivent être positives. Par conséquent, pour une répartition donnée φ_ν ($\nu = 1,2,\ldots,n$) et pour une loi de l'évolution donnée $k_{\nu\zeta}$ ($\nu,\zeta = 1,2,\ldots,n$), les proportions des masses des populations ne peuvent pas, ici non plus, être arbitraires. Les fréquences $f_{2\nu}$, de la population π_2 seront ensuite trouvées par les formules

$$f_{2\nu} = \frac{1}{q}(\varphi_\nu - pf_{1\nu})\,, \qquad (\nu = 1,2,\ldots,n)\,. \tag{13.8}$$

Exemple 13.1. - Soit la loi de Laplace

$$\varphi(x) = \frac{1}{a\sqrt{2\pi}}\,e^{-x^2/2a^2}$$

la loi de probabilité de la population totale π et supposons que la variable aléatoire x de la population π_2 soit obtenue au moyen de la variable aléatoire ξ de la population π_1 par la relation

$$x = \frac{t-\xi}{2}\,.$$

où t représente aussi une variable aléatoire avec la loi de probabilité

$$r(t) = \frac{1}{\sqrt{2\pi}}\,e^{-t^2/2}\,.$$

La loi de l'évolution aura la forme suivante

$$k(x,\xi) = \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}(2x+\xi)^2}$$

et la loi de probabilité itérée deviendra

$$\varphi_2(x) = \frac{1}{\pi a}\int_{-\infty}^{+\infty} e^{-\frac{\xi^2}{2a^2}-\frac{1}{2}(2x+\xi)^2}\, d\xi \quad =$$

$$= \frac{1}{\pi a} e^{-2x^2}\int_{-\infty}^{+\infty} e^{-\frac{1+a^2}{2a^2}\xi^2-2x\xi}\, d\xi \quad =$$

$$= \frac{2}{\sqrt{1+a^2}}\ \frac{1}{\sqrt{2\pi}} e^{-\frac{2x^2}{1+a^2}} .$$

De même

$$\varphi_n(x) = \frac{1}{\sigma_n\sqrt{2\pi}} e^{-x^2/2\sigma_n^2} ,$$

où l'écart-type σ_n est donné par la formule de récurrence suivante

$$\sigma_n^2 = \frac{1+\sigma_{n-1}^2}{4}, \quad \sigma_1^2 = a^2, \qquad n = 1, 2, \dots, \tag{13.9}$$

d'où

$$\sigma_n^2 = \frac{1}{3} + \frac{3a^2-1}{3 . 4^{n-1}} .$$

La suite $\{\sigma_n^2\}$ est monotone et tend vers 1/3 quand $n \to \infty$. Par conséquent, les lois de probabilité itérées seront Laplaciennes avec la moyenne arithmétique égale à zéro et les écarts-types variant entre a et $\sqrt{3}/3$. Elles ne peuvent pas dégénérer en fonctions de Dirac, donc nous pouvons utiliser la formule (13.5) et nous obtenons ainsi

$$f_1(x) = \frac{1}{p\sqrt{2\pi}}\sum_{\nu=0}^{\infty}\lambda^\nu\sqrt{\frac{3 . 4^\nu}{3a^2-1+4^\nu}}\, e^{-\frac{x^2}{2}\frac{3 . 4^\nu}{3a^2-1+4^\nu}} ,$$

$$f_2(x) = \frac{-1}{q\sqrt{2\pi}}\sum_{\nu=1}^{\infty}\lambda^\nu\sqrt{\frac{3 . 4^\nu}{3a^2-1+4^\nu}}\, e^{-\frac{x^2}{2}\frac{3 . 4^\nu}{3a^2-1+4^\nu}} .$$

Il reste encore à examiner la positivité des fonctions $f_1(x)$ et $f_2(x)$. Elles sont représentées par des séries alternées de fonctions. Or, elles seront certainement positives si l'inégalité

$$\frac{1}{\sigma_\nu\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_\nu^2}} \geqslant |\lambda|\ \frac{1}{\sigma_{\nu+1}\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_{\nu+1}^2}}$$

pour chaque ν et pour chaque x sur $(-\infty, +\infty)$. En utilisant la formule de récurrence (13.9), cette inégalité se réduit à

$$|\lambda| \leqslant \frac{\sqrt{1+\sigma_\nu^2}}{2\sigma_\nu} e^{-\frac{x^2}{2}\frac{3\sigma_\nu^2-1}{\sigma_\nu^2(1+\sigma_\nu^2)}} .$$

Considérons le cas où $\sigma_\nu^2 \geqslant 1/3$. L'inégalité ci-dessus sera satisfaite à plus forte raison si

$$|\lambda| \leqslant \frac{\sqrt{1+\sigma_\nu^2}}{2\sigma_\nu} ,$$

où

$$\frac{1-p}{p} \leqslant \frac{\sqrt{1+\sigma_\nu^2}}{2\sigma_\nu} ,$$

d'où

$$p \geqslant \frac{2\sigma_\nu}{2\sigma_\nu + \sqrt{1+\sigma_\nu^2}} . \tag{13.10}$$

L'expression figurant dans le deuxième membre de cette inégalité est une fonction de σ_ν qui croit de 0 à 2/3 pendant que σ_ν croit de 0 à $+\infty$. Etant donné que dans notre cas σ_ν varie entre $\sqrt{3}/3$ et a, l'inégalité (13.10) sera satisfaite pour chaque ν si

$$p \geqslant \frac{2a}{2a + \sqrt{1+a^2}} .$$

De $\sigma_\nu^2 \geqslant 1/3$, on en déduit que $a \geqslant \sqrt{3/3}$.

Les moyennes arithmétiques des populations π_1 et π_2 sont $m_1 = m_2 = 0$ et les variances

$$\Sigma_1^2 = \frac{1}{p}\sum_{\nu=0}^{\infty} \lambda^\nu \frac{3a^2-1+4}{3.4}$$

$$\Sigma_2^2 = -\frac{1}{q}\sum_{\nu=1}^{\infty} \lambda^\nu \frac{3a^2-1+4}{3.4}$$

Exemple 13.11. - La mortalité infantile est encore toujours très élevée en dépit du fait qu'elle est en baisse régulière dans la plupart des pays grâce aux mesures de protection dont jouissent chaque jour davantage les nourissons et les futures mères. On peut diviser les décès en deux catégorie principales. Dans la première catégorie se classent les décès des enfants nés en bonne santé. Leurs fréquences sont basses au cours des premiers mois de l'existence pour devenir ensuite de plus en plus élevées au détriment de la deuxième catégorie où la mortalité est favorisée par des causes héréditaires ; débilité à la naissance, prématuration, sous-alimentation maternelle ; hérédité constitutionnelle, syphilis, tuberculose. Les fréquences de cette deuxième catégorie sont très élevées au cours des quatre premières semaines suivant la naissance. Le Tableau 13.1 donne la répartition des enfants décédés au cours de la première année de vie en Suisse (1951) (38) d'après la durée de leur existence

TABLEAU 13.1 ENFANTS DECEDES AU COURS DE LA PREMIERE ANNEE DE VIE

Ages (en mois)	1-3	4-6	7-9	10-12	Total
Nombres absolus	2019	220	134	94	2467
En pourcentage	81,840	8,918	5,432	3,810	100

Ce Tableau donne les fréquences φ_ν (ν = 1,2,3,4) de la population totale π composée des populations décrites ci-dessus, π_1 et π_2. Dans le cas où les fréquences des populations π_1 et π_2 des années précédentes sont connues, nous avons vu dans le paragraphe II comment il est possible de déterminer les coefficients $K_{\nu\zeta}$ qui permettent de résoudre le système (13.7) lorsque nous connaissons les valeurs p et q. Supposons que p = 20 % et q = 80 % et essayons de reconstruire les répartitions de fréquences des populations π_1 et π_2 sans connaitre leurs répartitions antérieures.

De par la nature des deux populations nous pouvons supposer que π_1 se transforme en π_2 de telle façon que la durée de l'existence ξ de la population π_2 s'exprime au moyen de la durée de l'existence x de la population π_1 par la relation

$$x = t\xi \quad , \quad t \geqslant 0 ,$$

où la probabilité du facteur de proportionnalité t doit être d'autant plus grande qu'il est plus petit. C'est pourquoi nous pouvons prendre

$$\rho(t) = a\, e^{-at} ,$$

où a est un paramètre pour le moment, inconnu. Les coefficients $K_{\nu\zeta}$ (ν, ζ = 1,2,3,4) deviendront

$$k_{\nu\zeta} = a \int_{(\nu-1)h}^{\nu h} \frac{1}{\xi_\zeta} e^{-ax/\xi_\zeta}\, dx = \int_{ah(\nu-1)/\xi_\zeta}^{ah\nu/\xi_\zeta} e^{-t}\, dt.$$

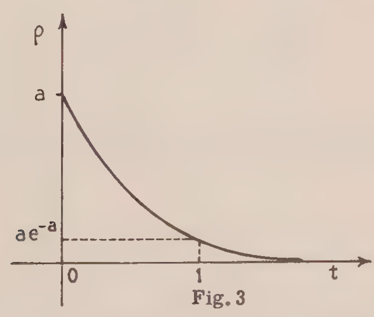


Fig. 3

En prenant ξ_ζ comme point médian du ζ -ième intervalle, c'est-à-dire

$$\frac{h}{\xi_\zeta} = 2/(2\zeta - 1)$$

il vient

$$k_{\nu\zeta} = e^{-\frac{2(\nu-1)}{2\zeta-1}a} - e^{-\frac{2\nu}{2\zeta-1}a}, \quad \nu, \zeta = 1,2,3,4.$$

Les conditions

$$k_{\nu\zeta} \geqslant 0, \qquad \nu, \zeta = 1,2,3,4,$$

$$\sum_{\nu=1}^{4} k_{\nu\zeta} = 1, \qquad \zeta = 1,2,3,4,$$

doivent être satisfaites, La première est toujours satisfaite. D'autre part, étant donné que

$$\sum_{\nu=1}^{4} k_{\nu\zeta} = 1 - e^{-\frac{8a}{2\zeta-1}}, \qquad \zeta = 1,2,3,4,$$

il est nécessaire que le paramètre a soit choisi de telle façon que l'erreur maximale exp. (-8a/ 7) puisse être négligée. Pour a = 4, cette erreur s'élève à 1 % et les valeurs des coefficients $k_{\nu\zeta}$ correspondantes sont données dans le Tableau 13.2

TABLEAU 13.2

	1	2	3	4
1	0,99966	0,93075	0,79810	0,68018
2	0,00034	0,06426	0,16114	0,21754
3	0,00000	0,00465	0,03253	0,07053
4	0,00000	0,00034	0,00823	0,03175

La solution du système

$$\begin{bmatrix} 1,250 & 0,931 & 0,798 & 0,680 \\ 0 & 0,314 & 0,161 & 0,218 \\ 0 & 0 & 0,283 & 0,070 \\ 0 & 0 & 0,008 & 0,282 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{14} \end{bmatrix} = \begin{bmatrix} 1,023 \\ 0,111 \\ 0,068 \\ 0,048 \end{bmatrix}$$

représentera la répartition de fréquences de la population π_1 . Le déterminant du système est Δ = 0,031104 et les fréquences relatives et absolues des populations π_1 et π_2 sont données dans les Tableaux 13.3 et 13.4 respectivement.

TABLEAU 13.3

$f_{i\nu}$ \ ν	1	2	3	4
$f_{1\nu}$	0,49906	0,13692	0,19950	0,16452
$f_{2\nu}$	0,89814	0,07700	0,01828	0,00658

TABLEAU 13.4

ν	1	2	3	4	Total
$N_1 f_{1\nu}$	246	68	98	81	493
$N_2 f_{2\nu}$	1.773	152	36	13	1.974
$N \varphi_\nu$	2.019	220	134	94	2.467

RÉFÉRENCES

(1) J. WISHART (1928), "The generalized product moment distribution in samples from a normal multivariate population", Biometrika, Vol 20 A, p. 32

(2) M.S. BARLETT et J. WISHART (1932), "The generalized product moment distribution in a normal system", Proc. Cambridge Phil. Soc., Vol 29, p. 260.

(3) H. HOTELLING (1931), "The generalization of Student's ratio", Annals Of the Math. Statistics, Vol. 2, p. 360.

(4) "STUDENT" (1908), "The probable error of a mean", Biometrika, Vol. 6, p. 1.

(5) M.S. BARTLETT (1938), "Further aspects of the theory of multiple regression", Proc. Cambridge Phil. Soc. Vol. 34, p. 33

(6) S.S. WILKS (1935), "On the indenpendence of k sets of normally distributed statistical variables", Econometrica, Vol.3 p.309

(7) M.S.BARTLETT (1934), "The vector représentation of a sample", Proc. Cambridge Phil. Soc., Vol. 30, p. 327

(8) H. HOTELLING (1935), "The relation between two sets of variates", Biometrika, Vol. 28, p. 328.

(9) R.L. PLACKETT (1947), "An exact test for the equality of variances", Biometrika, Vol. 34, p. 311.

(10) C.R. RAO (1946), "Tests with discriminant functions in multivariate analysis", Sankhya, Vol. 7, p. 407.

(11) C.R. RAO, "Tests of significance in multivariate analysis", Biometrika, Vol. 35, p. 58. (1948)

(12) R.A. FISHER (1936), "The use of multiple measurements in taxonomic problems", Annals of Eugenics, London, Vol, 7, p. 240.

(13) R.A. FISHER (1938), "The statistical utilization of multiple measurements", Ann. Eugen., London, Vol. 8, p. 376

(14) R.A. FISHER (1940), "The precision of the discriminant function", Ann. Eugen. London, Vol. 10, p. 422

(15) R.A.FISHER (1939), "The sampling distribution of some statistics obtained from non linear regression", Ann. Eugen., London, Vol. 9, p. 238.

(16) M.M.BARNARD (1935), "The secular variation of skull characters in four series of Egyptian skulls", Ann. Eugen., London, Vol. 7, p. 89.

(17) E.S.MARTIN (1935), "A study of the Egyptian series of mandiables, with special reference to mathematical methods of sexing", Biometrika, Vol. 28, p. 149.

(18) FAIRFIELD SMITH (1936), "A discriminant function for plant selection", Ann. Eugen., London Vol. 7, p. 240.

(19) G.M.COX et W.P.MARTIN (1937), "Use of a discriminant function for differentiating Soils with different Azotobacter populations", Yowa State Col. Journ. Sci. Vol. 9, p. 323.

(20) D.S. PENROSE(1945), "Discrimination between normal and psychotic subjects by examination M", Bull. Canad. Psychol. Ass. Vol. 5, p. 37.

(21) D.DURANT (1941), "Risk elements in consumer instalment Financing", Nath. Bur. Econ. Res. Inc., Studies in Consumer Instalment Financing, N° 8.

(22) P.C.MAHALONOBIS (1927), "Analysis of race mixture in Bengal", Journ. Asiat. Soc. Beng., Vol. 23, p. 301.

(23) P.C.MANALANOBIS (1930), "On tests and measures of group divergence", Journ. Asiat. Soc. Beng., Vol. 26, p. 541.

(24) N.WALACE et R.M.W.TRAVERS (1938), "A psychometric sociological study of a group of speciality Solesmen", Ann.Eugen. London, Vol. 8, p. 266.

(25) B.B.DAY et M.M.SANDOMIRE (1942), "Use of the discriminant function for more than two groups", Journ. Amer. Statist. Assoc., Vol. 37, p. 461.

(26) P.C.MAHALANOBIS (1936), "On the generalized distance in statistics", Proc. Nat. Inst. Sci. (India), Vol. 12, p. 49.

(27) C.R.RAO (1948), "The utilization of multiple measurements in problems of biological classification", Journ. Roy. Stat. Soc, Vol. 10, p. 159.

(28) B.D. WELSH (1939), "Note on discriminant functions", Biometrika, Vol. 31, p. 218.

(29) C. A. B. SMITH (1947), "Some examples of discrimination", Ann. Eugen., London, Vol. 13, p. 272.

(30) G.M.MORANT (1926), "A first study of craniology of England and Scotland from Meolithic to early historic times, with special reference to Anglo-Saxon skulls in London Museums", Biométrika, Vol. 18, p. 36.

(31) C.N.GOODWIN et G.M.MORANT (1940), "The humain remains of Iron Age and other periods from Maiden Castle, Dorset, Biometrika, Vol. 31, p. 295.

(32) B.G.E.HOOKE (1926), "A third study of the English skull with special reference to the Farrington Street crania", Biometrika, Vol. 18, p. 1

(33) P.C.MAHALONOBIS. D.N.MAJUMDAR et C.R.RAO (1949), "Anthropometric survey of the United Provinces, 1941 ; A Statistical Study", Sankhya, Vol. 9, p. 90.

(34) C.R.RAO (1950), "Statistical inference applied to classificatory problems", Sankhya, Vol. 10, p. 229.

(35) H.CRAMER (1937), "Random Variables and Probability Distributions" Cambridge Tracts in Mathematics, n° 36, p. 46.

(36) C.A. BAUD et M.J. DALLEMAGNE (1949), "A new method for the study of submicroscopic spaces", Science (N.Y.), Vol.110 p. 90-91.

(37) K.PEARSON (1894), "Contributions to the mathematical theory of evolution. I. Dissection of frequency curves", Philos. Trans., A. 185, p. 71.

(38) "Annuaire statistique de la Suisse", Birkhäuser, Bâle (1952).

(38) "Statistisches Jahrbuch des Kantons Basel-Stadt", Bâle (1953).

TABLE DES MATIÈRES

TABLE DES MATIÈRES DU VOLUME III

ANNÉE 1954

J. & R. SENNAC, Imprimeurs
54, Fg Montmartre, PARIS (9e)